

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

M protein - protein search, using sw model

un on: August 6, 2002, 17:05:14 : Search time 66.51 seconds
(without alignments)
24.560 Million cell updates/sec

title: US-10-020-139-2_COPY_68_84

effect score: 83

sequence: 1 SSAWLAQKQAQAEKL 17

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 283138 seqs, 96089334 residues

total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

database:

PIR_71:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	48	57.8	956	2 B71250	valine--trna ligase
2	46	55.4	858	2 D71711	endopeptidase Clp
3	43	51.8	270	2 A80514	DnaJ-like protein
4	43	51.8	685	2 G82022	methionine--trna l
5	42	50.6	736	2 E71414	hypothetical prote
6	42	50.6	844	1 A26528	penicillin amidase
7	42	50.6	846	1 PNECA	penicillin amidase
8	41	49.4	185	2 D70437	ATP synthase FO su
9	41	49.4	188	2 S32862	outH protein - Erw
10	41	49.4	334	2 A10795	probable receptor/
11	41	49.4	409	2 T46922	hypothetical prote
12	41	49.4	844	2 AC1981	hypothetical prote
13	41	49.4	857	2 C97707	clpB protein [impo
14	41	49.4	995	2 S67564	hypothetical prote
15	41	49.4	1148	2 D83270	transcriptional repa
16	41	49.4	1232	2 S40766	hypothetical prote
17	41	49.4	1952	2 T48814	hypothetical prote
18	40	48.2	125	2 A81644	protein gp30 (bact
19	40	48.2	125	2 A11358	protein gp30 (bact
20	40	48.2	158	2 T46199	hypothetical prote
21	40	48.2	171	2 AD0309	conserved hypothe
22	40	48.2	190	2 A84029	septum formation m
23	40	48.2	254	2 S75187	hypothetical prote
24	40	48.2	262	2 T48221	divIB protein - Ba
25	40	48.2	341	2 T18873	hypothetical prote
26	40	48.2	396	2 E87947	cell division prot
27	40	48.2	399	2 B95080	cell division prot
28	40	48.2	412	2 AB1000	type II secretion
29	40	48.2	469	2 T48783	related to SGT1 pr

30 40 48.2 578 2 AG0119 probable ompa-fami
31 40 48.2 578 2 AC0179 probable exported
32 40 48.2 629 2 AG1799 GidA protein homol
33 40 48.2 632 2 T02627 hypothetical prote
34 40 48.2 715 2 G75339 probable exodeoxyr
35 40 48.2 733 2 T23507 hypothetical prote
36 40 48.2 785 2 B84038 DNA mismatch repai
37 40 48.2 1093 2 B86748 hypothetical prote
38 40 48.2 1254 2 A54818 myosin-V [similar
39 40 48.2 1265 2 A59299 unconventional myo
40 40 48.2 1296 2 E81840 transcriptional repa
41 40 48.2 1379 2 A81102 transcriptional repa
42 40 48.2 2022 2 T48818 glucan 1,4-alpha-g
43 39.5 47.6 439 2 S67921 multiple ligand-bi
44 39.5 47.6 685 2 A81246 methionyl-trna syn
45 39.5 47.6 907 2 D83427 probable transcrip

ALIGNMENTS

RESULT 1

B71250

valine--trna ligase (EC 6.1.1.9) (vals) - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 24-Nov-1999

C:Accession: B71250

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.O.; Dodson, R.; rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770

A:Accession: B71250

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-956 <COL>

A:Cross-references: GB:AE001271; GB:AE000520; NID:g3323363; PIDN:AAC26589.1; PID:g33

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TPI035

C:Superfamily: valine--trna ligase

C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 57.8%; Score 48; DB 2; Length 956;
Best Local Similarity 57.1%; Pred. No. 14;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 AWQLAKOKAQAQAEK 16

III I:II I:II

Db 892 AWQKARQKQVQVER 905

RESULT 2

D71711

endopeptidase Clp ATP-binding chain B - Rickettsia prowazekii

N:Alternate names: ATP-dependent Clp proteinase regulatory chain

N:Contains: adenosinetriphosphatase (EC 3.6.1.3)

C:Species: Rickettsia prowazekii

C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 19-Jan-2001

C:Accession: D71711

R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmar Nature 396, 133-140, 1998

A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria

A:Reference number: A71630; MUID:99039499

A:Accession: D71711

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-838 <AND>

A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14507.1; PID:g38

A:Experimental source: strain Madrid E

C:Genetics:

A:gene: both
 C:Superfamily: secretion protein xcpU
 Query Match 49.4%; Score 41; DB 2; Length 188;
 Best Local Similarity 72.7%; Pred. No. 36;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

2 SAWQLAKQAK 12
|||||: |||
42 SAWQLARFOAQ 52

SULT 10

0795
obable receptor/regulator protein STY2545 [imported] - Salmonella enterica subsp. ente
Species: Salmonella enterica subsp. enterica serovar Typhi
Note: this species has also been called Salmonella typhi
Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
Accession: A10795
Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Garra, P.
ture 413, 848-852, 2001
Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
Reference number: AB0502; PMID:11677608
Accession: A10795
Status: preliminary
Molecule type: DNA
Residues: 1-334 <PAR>
Cross-references: GB:AL513382; PIDN:CAD07547.1; PID:g16503539; GSPDB:GN00176
Genetics:
Gene: STY2545
Superfamily: chemotaxis chev protein; response regulator homology

Query Match 49.4%; Score 41; DB 2; Length 334;
Best Local Similarity 61.5%; Pred. No. 64;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

3 AWQLAKQAKQAE 15
||: ||| |||
222 AWERIQQAQAE 234

SULT 11

6922
pothetical protein DKFZp434B1612.1 - human
Species: Homo sapiens (man)
Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 17-Mar-2000
Accession: T46922
Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Weil, B.; Wiemann, S.
mitted to the Protein Sequence Database, February 2000
Reference number: Z24138
Accession: T46922
Status: preliminary
Molecule type: mRNA
Residues: 1-409 <RAA>
Cross-references: EMBL:AL157436
Experimental source: adult testis; clone DKFZp434B1612
Genetics:
Note: DKFZp434B1612.1

Query Match 49.4%; Score 41; DB 2; Length 409;
Best Local Similarity 50.0%; Pred. No. 78;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

2 SAWQLAKQAKQAE 17
|||: ||| | : ||
222 SAWELGKVKGLKEEL 237

SULT 12

1981
pothetical protein alr1398 [imported] - Anabaena sp. (strain PCC 7120)
Species: Anabaena sp.
Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
Accession: AC1981

R: Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A: Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A: Reference number: AB1807; MUID:21595285; PMID:11759840
A: Accession: AC1981
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-844 <KUR>
A: Cross-references: GB:BA000019; PIDN:BA073355.1; PID:g17130745; GSPDB:GN00179
A: Experimental source: strain PCC 7120
C: Genetics:
A: Gene: alr1398

Query Match 49.4%; Score 41; DB 2; Length 844;
Best Local Similarity 58.3%; Pred. No. 1.6e+02;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 AWQLAKQAKQAE 14
: ||||: |||:
Db 128 SWQLAQKQAEQA 139

RESULT 13

C97707
clpB protein [imported] - Rickettsia conorii (strain Malish 7)
C: Species: Rickettsia conorii
C: Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C: Accession: C97707
R: Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.;
Science 293, 2093-2098, 2001
A: Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A: Reference number: A97700; MUID:21442074; PMID:11557893
A: Accession: C97707
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-857 <KUR>
A: Cross-references: GB:AE006914; PIDN:AAL02597.1; PID:g15619095; GSPDB:GN00173
C: Genetics:
A: Gene: clpB
C: Superfamily: endopeptidase Clp ATP-binding chain

Query Match 49.4%; Score 41; DB 2; Length 857;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 WQLAKQAKQAEK 17
|| | | | : ||
Db 463 WQAEKSKLQQTQKL 476

RESULT 14

S67564
hypothetical protein YDL031w - yeast (Saccharomyces cerevisiae)
A: Alternate names: hypothetical protein D2770
C: Species: Saccharomyces cerevisiae
C: Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 02-Feb-2001
C: Accession: S67564
R: Paulin, L.; Saren, A.M.; Laamanen, P.
submitted to the Protein Sequence Database, July 1996
A: Reference number: S67560
A: Accession: S67564
A: Molecule type: DNA
A: Residues: 1-995 <PAU>
A: Cross-references: EMBL:Z74079; MID:g1431009; PIDN:CAA98590.1; PID:g252986; PID:g143
A: Experimental source: strain S288C
C: Genetics:
A: Map position: 4L
C: Keywords: ATP; nucleotide binding; P-loop
F: 181-188/Region: nucleotide-binding motif A (P-loop)
F: 284-289/Region: nucleotide-binding motif B

;288-291/Region: DEAD motif

Query Match 49.4%; Score 41; DB 2; Length 995;
 Best Local Similarity 69.2%; Pred. No. 1.9e+02;
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

y 5 QLAKQXAQEAEXL 17
 ||||| : |||||
 b 621 QLAKKTAMKAEXL 633

RESULT 15

83270
 ;transcription-repair coupling protein Mid PA3002 [imported] - Pseudomonas aeruginosa (su
 ;Species: Pseudomonas aeruginosa
 ;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 ;Accession: D83270
 ;Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
 ;dhan, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
 ;Lory, S.; Olson, M.V.
 ;ature 406, 959-964, 2000
 ;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 ;Reference number: AB2950; MUID:20437337
 ;Accession: D83270
 ;Status: preliminary
 ;Molecule type: DNA
 ;Residues: 1-1148 <STO>
 ;Cross-references: GB:AE004725; GB:AE004091; NID:99949100; PIDN:AG06390.1; GSPDB:GN001
 ;Experimental source: strain PA01
 ;Genetics:
 ;Gene: mfd; PA3002
 ;Superfamily: transcription-repair coupling protein

Query Match 49.4%; Score 41; DB 2; Length 1148;
 Best Local Similarity 61.5%; Pred. No. 2.2e+02;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

y 1 SSAWQLAKOKAQE 13
 | | | | | | | |
 b 548 SETWQAKRKAQE 560

earch completed: August 6, 2002, 17:05:15
 ob time: 917 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

DM protein - protein search, using sw model

Run on: August 6, 2002, 17:07:34 ; Search time 32.88 Seconds

(without alignments)

20.019 Million cell updates/sec

Title: US-10-020-139-2_COPY_68_84

Perfect score: 83

Sequence: 1 SSAWQLAKKQAEKEL 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	57.8	956	1 SVV_TREPA	O83998 treponema p
2	42	50.6	844	1 PAC_KLUCI	P07941 kluyvera ci
3	42	50.6	846	1 PAC_ECOLI	P06875 escherichia
4	41	49.4	104	1 RPOZ_STRPY	P82577 streptococc
5	41	49.4	188	1 GSPH_ERWCA	P31587 erwinia car
6	41	49.4	995	1 DBPA_YEAST	Q12389 saccharomyc
7	41	49.4	1232	1 YQO5_CAEEL	P34643 caenorhabdi
8	40	48.2	190	1 YAF_BACHD	Q9K8H3 bacillus ha
9	40	48.2	341	1 YKLI_CAEEL	P42168 caenorhabdi
10	40	48.2	785	1 MUS2_BACHD	Q9K8A0 bacillus ha
11	40	48.2	1262	1 MYO6_HUMAN	Q9UM54 homo sapien
12	40	48.2	1265	1 MYO6_MOUSE	O64331 mus musculu
13	39	47.0	181	1 GSPH_ERWCH	P24687 erwinia chr
14	39	47.0	195	1 AANT_HDVS1	P25853 hepatitis d
15	39	47.0	195	1 AANT_HDVS2	P25854 hepatitis d
16	39	47.0	271	1 DULA_ECOLI	P1680 escherichia
17	39	47.0	544	1 THS_METTL	O93624 methanococc
18	39	47.0	575	1 YFBK_ECOLI	P76481 escherichia
19	39	47.0	1750	1 Y832_METJA	Q58242 methanococc
20	38	45.8	180	1 CEAS_ECOLI	P18000 escherichia
21	38	45.8	197	1 RNH2_PASMU	P57996 pasteurella
22	38	45.8	205	1 CEAB_ECOLI	O99852 escherichia
23	38	45.8	280	1 T2M1_MORBO	P34719 moraxella b
24	38	45.8	379	1 METX_CORGL	O68640 corynebacte
25	38	45.8	462	1 SYC_CAMAE	O99P88 campylobact
26	38	45.8	505	1 NDHF_BACSU	P19755 bacillus su
27	38	45.8	532	1 MURD_TREPA	O83873 treponema p
28	38	45.8	536	1 SP01_YEAST	P53541 saccharomyc
29	38	45.8	537	1 P2V8_XENLA	P79928 xenopus lae
30	38	45.8	551	1 CEAB_ECOLI	P00646 escherichia
31	38	45.8	551	1 CEAB_ECOLI	P17999 escherichia
32	38	45.8	576	1 CEAB_ECOLI	O47112 escherichia
33	38	45.8	581	1 CEAB_ECOLI	P04419 escherichia

34	38	45.8	582	1 CEAB_ECOLI	P09883 escherichia
35	38	45.8	682	1 SYM_HAEIN	P43828 haemophilus
36	38	45.8	682	1 SYM_PASMU	P57838 pasteurella
37	38	45.8	860	1 MUTS_PASMU	P57972 pasteurella
38	38	45.8	919	1 RA50_AERPE	Q9YF21 aeropyrum p
39	38	45.8	1054	1 SPSI_CRAPL	O04932 craterostig
40	38	45.8	1145	1 Y623_DROME	O9VBT7 drosophila
41	38	45.8	1532	1 IGA_NEIGO	P09790 neisseria g
42	38	45.8	1612	1 TP2B_CRILO	O64399 cricetulus
43	38	45.8	1612	1 TP2B_MOUSE	O64511 mus musculu
44	38	45.8	1626	1 TP2B_HUMAN	O02880 homo sapien
45	38	45.8	1627	1 TP2B_CHICK	O42131 gallus gall

ALIGNMENTS

RESULT 1

SVV_TREPA
ID SVV_TREPA STANDARD: PRT; 956 AA.
AC O83998;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Valyl-tRNA synthetase (EC 6.1.1.9) (Valine--tRNA ligase) (VALRS).
GN VALS OR TP1035.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
spirochete."
RL Science 281:375-388(1998).
CC -!- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate
+ L-valyl-tRNA(Val).
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: AE001271; AAC36589.1;
CC TIGR: TP1035;
DR InterPro: IPR002300; tRNA-synt_la.
DR InterPro: IPR001412; tRNA-synt_1.
DR InterPro: IPR002303; tRNA-synt_val.
DR Pfam: PF00133; tRNA-synt_1; 1.
DR PRINTS: P00986; TRNASYNTHAL.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
Complete proteome.
FT SITE 69 79 "HIGH" REGION.
FT SITE 566 570 "KMSKS" REGION.
FT BINDING 569 569 ATP (BY SIMILARITY).
FT SEQUENCE 956 AA; 107362 MW; 7D50583C7C6BA35A CRC64;
SQ

Query Match 57.8%; Score 48; DB 1; Length 956;
 Best Local Similarity 57.1%; Pred. No. 6.2;
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Y 3 AWOLAKOQAEK 16

|||||I|I|I|I|

b 892 AWQKQKVQOVER 905

RESULT 2

AC_KLUCCI STANDARD; PRT; 844 AA.
 D PAC_KLUCCI
 C P07941;
 T 01-JAN-1988 (Rel. 08, Created)
 T 01-AUG-1988 (Rel. 08, Last sequence update)
 T 16-OCT-2001 (Rel. 40, Last annotation update)
 T Penicillin G acylase precursor (EC 3.5.1.11) (Penicillin G amidase)
 T (Penicillin amidohydrolase).
 N PAC.
 S Kluyvera citrophila (Kluyvera cryocrescens).
 C Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 C Kluyvera.
 X NCBI_TaxID=580;
 N [1]
 P SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 C STRAIN=ATCC 21285;
 C MEDLINE=87192002; PubMed=3032748;
 A Barbero J.L., Buesa J.M., de Buitrago G.G., Mendez E.,
 A Perez-Aranda A., Garcia J.L.;
 T "Complete nucleotide sequence of the penicillin acylase gene from
 T Kluyvera citrophila";
 L N Gene 49:69-80(1986).
 N [2]
 P SEQUENCE OF 290-297, AND ACTIVE SITE.
 X MEDLINE=92109664; PubMed=1764029;
 A Martin J., Slade A., Aiken A., Arche R., Virden R.;
 T "Chemical modification of serine at the active site of penicillin
 T acylase from Kluyvera citrophila";
 L Biochem. J. 280:659-662(1991).
 C -1- CATALYTIC ACTIVITY: Penicillin + H(2)O -> a fatty acid anion + 6-
 C aminopenicillanate.
 C -1- SUBUNIT: HETERODIMER OF AN ALPHA CHAIN AND A BETA CHAIN.
 C -1- SUBCELLULAR LOCATION: Periplasmic.
 C -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S45 (SERINE PROTEASE) ALSO
 C KNOWN AS THE PENICILLIN ACYLASE FAMILY.

 C This SWISS-PROT entry is copyright. It is produced through a collaboration
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -
 C the European Bioinformatics Institute. There are no restrictions on its
 C use by non-profit institutions as long as its content is in no way
 C modified and this statement is not removed. Usage by and for commercial
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 C or send an email to license@sib-sib.ch).
 C -----

R EMBL; M15418; AAA25047.1;
 R PIR; A26528; A26528.
 R HSP; P06875; 1AJQ.
 R MEROPS; S45.001; -

R InterPro: IPR002692; Penicil_amidase.

R Pfam: PF01804; Penicil_amidase; 1.

R W Hydrolyase; Antibiotic resistance; Zymogen; Periplasmic; Signal.

T SIGNAL

T CHAIN 1 26

T CHAIN 27 844

T CHAIN 235

T PROPEP 236 289

T CHAIN 290 844

T ACT_SITE 290 290

T SEQUENCE 844 AA; 93572 MW; 3466059B0FC4E5AF CRC64;

Query Match 50.6%; Score 42; DB 1; Length 844;
 Best Local Similarity 41.2%; Pred. No. 47;

Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 SSAWQLAKOQAEK 17

|||||I|I|I|I|

Db 290 SNMWVICKNKAQDAK 306

RESULT 3

PAC_ECOLI
 ID PAC_ECOLI STANDARD; PRT; 846 AA.
 AC P06875;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Penicillin G acylase precursor (EC 3.5.1.11) (Penicillin G amidase)
 DE (Penicillin G amidohydrolase).
 GN PAC.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86286584; PubMed=3016663;
 RX Schumacher G., Sizmann D., Haug H., Buckel P., Boeck A.;
 RT "Penicillin acylase from E. coli: unique gene-protein relation.";
 RL Nucleic Acids Res. 14:5713-5727(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88056318; PubMed=3315861;
 RX Oh S.-J., Kim Y.-C., Park Y.-W., Min S.-Y., Kim I.-S., Kang H.-S.;
 RT "Complete nucleotide sequence of the penicillin G acylase gene and
 RT the flanking regions, and its expression in Escherichia coli";
 RL Gene 56:87-97(1987).
 RN [3]
 RP SEQUENCE OF 1-368 FROM N.A.
 RC STRAIN=ATCC 11105;
 RX MEDLINE=86137424; PubMed=3005131;
 RX Oliver G., Valle F., Rosetti F., Gomez-Pedrozo M., Santamaria P.,
 RA Gosset G., Bolivar F.;
 RT "A common precursor for the two subunits of the penicillin acylase
 RT from Escherichia coli ATCC11105";
 RL Gene 40:9-14(1985).
 RN [4]
 RP SEQUENCE OF 1-368 FROM N.A.
 RX MEDLINE=87219870; PubMed=3556321;
 RX Valle F., Gosset G., Tenorio B., Oliver G., Bolivar F.;
 RT "Characterization of the regulatory region of the Escherichia coli
 RT penicillin acylase structural gene";
 RL Gene 50:119-122(1986).
 RN [5]
 RP SEQUENCE OF 1-95 FROM N.A.
 RC STRAIN=ATCC 11105;
 RX Radoja S., Francetic O., Stojicevic N., Moric I., Glisin S.,
 RA Konstantinovic M.;
 RT "Transcriptional and gene fusion analyses of the Escherichia coli
 RT penicillin amidase gene expression";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP NATURATION.
 RX MEDLINE=90382435; PubMed=2205499;
 RX Sizmann D., Keilmann C., Boeck A.;
 RT "Primary structure requirements for the maturation in vivo of
 RT penicillin acylase from Escherichia coli ATCC 11105";
 RL Eur. J. Biochem. 192:143-151(1990).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
 RC STRAIN=ATCC 11105;
 RX MEDLINE=95115804; PubMed=7816145;
 RX Duggleby H.J., Tolley S.P., Hill C.P., Dodson E.J., Dodson G.,
 RA Moody P.C.E.;
 RT "Penicillin acylase has a single-amino-acid catalytic centre";
 RL Nature 373:264-268(1995).

SA Salmund G.P.C.:
 XT "Molecular cloning and characterization of 13 out genes from Erwinia
 XT carotovora subspecies carotovora: genes encoding members of a general
 XT secretion pathway (GSP) widespread in Gram-negative bacteria.";
 XL Mol. Microbiol. 8:443-456(1993).
 XC -1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
 XC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF THE
 XC MULTIPLE PEPTIC ENZYMES.
 CC -1- SIMILARITY: BELONGS TO THE PULV/OUTH/XPSH/EXEH/XCPU FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X70049; CAA49649.1;
 CC PIR: S31751; S31751.
 CC PIR: S32862; S32862.
 CC InterPro: IPR002416; Bac_GSPH.
 CC InterPro: IPR001120; Prok_N_methylLn.
 CC PRINTS: PR00885; BCTERIALGSPH.
 CC PROSITE: PS00409; PROKAR_NTER_METHYL; 1.
 CC Transport: Methylation.
 XT PROPEP 1 10 BY SIMILARITY.
 XT CHAIN 11 188 GENERAL SECRETION PATHWAY PROTEIN H.
 XT MOD_RES 11 11 METHYLATION (BY SIMILARITY).
 SQ SEQUENCE 188 AA; 21390 MW; 733E90BFD6DC5F9 CRC64;

Query Match 49.4%; Score 41; DB 1; Length 188;
 Best Local Similarity 72.7%; Pred. No. 16;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

2Y 2 SAWQIAKOKAQ 12
 IIILII: III
 42 SAWQLARFQAQ 52

RESULT 6
 DBPA_YEAST STANDARD; PRT; 995 AA.
 AC Q12389;
 XT 01-NOV-1997 (Rel. 35, Created)
 XT 01-NOV-1997 (Rel. 35, Last sequence update)
 XT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable RNA-dependent helicase DBP10 (DEAD-box protein 10).
 SN DBP10 OR YD031W OR D2770.
 SC Saccharomyces cerevisiae (Baker's yeast).
 XC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 XC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 XN NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RA Paulin L., Saren A.M.;
 RA Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP CHARACTERIZATION:
 RX MEDLINE=20330350; PubMed=10871363;
 RA Burger F., Daugeron M.-C., Linder P.;
 XT "Dbp10, a putative RNA helicase from Saccharomyces cerevisiae, is
 XT required for ribosome biogenesis.";
 XL Nucleic Acids Res. 28:2315-2323(2000).
 XC -1- FUNCTION: PUTATIVE ATP-DEPENDENT RNA HELICASE. PROBABLY INTERACTS
 XC WITH 60S RIBOSOMAL SUBUNIT. INVOLVED IN THE MATURATION OF 25S
 XC RIBOSOMAL RNA.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
 CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. HIGHLY
 CC SIMILAR TO S.POMBE SPAC31A2.107C.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z71781; CAA96458.1;
 CC SGB: S0002189; DBP10.
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR000629; DEAD_ATP_helicase.
 CC InterPro: IPR001650; Helicase_C.
 CC Pfam: PF00270; DEAD; 1.
 CC Pfam: PF00271; Helicase_C; 1.
 CC SMART: SM00487; DEXHC; 1.
 CC SMART: SM00490; HELICC; 1.
 CC PROSITE: PS00039; DEAD_ATP_HELICASE; 1.
 KW ATP-binding; RNA-binding; Helicase; Nuclear protein.
 FT NP_BIND 188 ATP (POTENTIAL).
 FT SITE 288 291 DEAD BOX.
 SQ SEQUENCE 995 AA; 113157 MW; 9AFA74D0B8B198A2 CRC64;

Query Match 49.4%; Score 41; DB 1; Length 995;
 Best Local Similarity 69.2%; Pred. No. 78;
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 5 QLAKOKAOEAKL 17
 IIILII: I:IIILII
 DB 621 QLAKKTANKAEKL 633

RESULT 7
 YOO5_CAEEL STANDARD; PRT; 1232 AA.
 AC P34643;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Hypothetical 134.9 kDa protein ZK512.5 in chromosome III.
 GN ZK512.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Ronfield J., Hurton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favellio A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
 RA Sims M., Smalton N., Smith A., Smith M., Sonnhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Woulman P.;
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC -----
CC EMBL: Z22177; CA80146.1;
CC PIR: S40766; S40766
CC WormPep: ZK512.5; CE00411.
CC KW Hypothetical protein.
CC SEQUENCE 1232 AA; 134923 MW; 6DFC35D664AA8D6A CRC64;

Query Match 49.4%; Score 41; DB 1; Length 1232;
Best Local Similarity 53.3%; Pred. No. 96;
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 3 AWLAKQKQAEAEKL 17
Db 326 AYEKARQEAASEKL 340
      |::|::| |::|
      |::|::| |::|

RESULT 8
MAF_BACHD STANDARD; PRT; 190 AA.
AC Q9K8H3;
CC 01-MAR-2002 (Rel. 41, Created)
CC 01-MAR-2002 (Rel. 41, Last sequence update)
CC 01-MAR-2002 (Rel. 41, Last annotation update)
CC DE Septum formation protein Maf.
CC GN MAF OR BH3033.
CC OS Bacillus halodurans.
CC OC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC OC Bacillus/Staphylococcus group; Bacillus.
CC OX NCBI_TaxID=86665;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=C-125 / JCM 9153;
CC MEDLINE=20512582; PubMed=11058132;
CC RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
CC Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
CC Horikoshi K.;
CC "Complete genome sequence of the alkaliphilic bacterium Bacillus
CC halodurans" and genomic sequence comparison with Bacillus subtilis."
CC Nucleic Acids Res. 28:4317-4331(2000).
CC -!- FUNCTION: Involved in septum formation (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE MAF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP001517; BAB06752.1;
CC InterPro: IPR003697; Maf.
CC Pfam: PF02545; Maf; 1.
CC Complete proteome. 33
CC ACT_SITE 33
CC SEQUENCE 190 AA; 21240 MW; A810954296FBE099 CRC64;

Query Match 48.2%; Score 40; DB 1; Length 190;
Best Local Similarity 66.7%; Pred. No. 23;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 QLAKQKQAEAEK 16
Db 47 QLAKQKQADVAK 58
      |::|::| |
      |::|::| |

RESULT 9
YKLL_CAEEL STANDARD; PRT; 341 AA.
ID YKLL_CAEEL

```

```

AC P42168;
CC 01-NOV-1995 (Rel. 32, Created)
CC 01-NOV-1995 (Rel. 32, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Putative casein kinase I C03C10.1 in chromosome III (EC 2.7.1.1.-).
CC GN C03C10.1.
CC OS Caenorhabditis elegans.
CC OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
CC OC Rhabditidae; Peloderinae; Caenorhabditis.
CC OX NCBI_TaxID=6239;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=BRISTOL N2;
CC RA Berks M.;
CC Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN -> ADP + A PHOSPHOPROTEIN.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -!- CASEIN KINASE I SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z35637; CA84685.1;
CC HSSP: Q06486; ICKI.
CC WormPep: C03C10.1; CE00872.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002290; Ser_thr_pkinase.
CC Pfam: PF00069; pkinase; 1.
CC PROSITE: P500107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: P500108; PROTEIN_KINASE_ST; 1.
CC PROSITE: P500111; PROTEIN_KINASE_DOM; 1.
CC KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
CC ATP-binding.
CC FT DOMAIN 16 284 PROTEIN KINASE.
CC NP_BIND 22 30 ATP (BY SIMILARITY).
CC FT BINDING 45 45 ATP (BY SIMILARITY).
CC FT ACT_SITE 135 135 BY SIMILARITY.
CC SEQUENCE 341 AA; 39037 MW; 17AD6E173551BABC CRC64;

Query Match 48.2%; Score 40; DB 1; Length 341;
Best Local Similarity 50.0%; Pred. No. 40;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 WQLAKQKQAEAE 15
Db 297 WTMKQKQKQSQ 308
      |::|::|::|
      |::|::|::|

RESULT 10
MUS2_BACHD STANDARD; PRT; 785 AA.
ID MUS2_BACHD
CC Q9K8A0;
CC 01-MAR-2002 (Rel. 41, Created)
CC 01-MAR-2002 (Rel. 41, Last sequence update)
CC 01-MAR-2002 (Rel. 41, Last annotation update)
CC DE Muts2 protein.
CC GN MUTS2 OR BH3106.
CC OS Bacillus halodurans.
CC OC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC OC Bacillus/Staphylococcus group; Bacillus.
CC OX NCBI_TaxID=86665;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=C-125 / JCM 9153;
CC MEDLINE=20512582; PubMed=11058132;
CC RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
CC Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
CC -----

```


RESULT 12					
MYO6_MOUSE	MYO6_MOUSE	STANDARD;	PRT;	1265 AA.	
ID	MYO6_MOUSE				
AC	Q64331;				
DT	16-OCT-2001	(Rel. 40, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	01-MAR-2002	(Rel. 41, Last annotation update)			
DE	Myosin VI.				
GN	MYO6 OR SV.				
OS	Mus musculus (Mouse).				

```

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96083582; PubMed=7493015;
RA Abraham K.B., Hasson T., Steel K.P., Kingsley D.M., Russell L.B.,
RA Moosker M.S., Copeland N.G., Jenkins N.A.;
RT "The mouse Snell's waltzer deafness gene encodes an unconventional
RT myosin required for structural integrity of inner ear hair cells.";
RL Nat. Genet. 11:369-375(1995).
CC -!- FUNCTION: RECESSIVE ACTIN-BASED MOTOR. REQUIRED FOR STRUCTURAL
CC INTEGRITY OF INNER EAR HAIR CELLS (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. WITHIN THE COCHLEA,
CC EXPRESSED SPECIFICALLY WITHIN THE SENSORY HAIR CELLS.
CC -!- DISEASE: DEFECTS IN MYO6 ARE THE CAUSE OF SNELL'S WALTZER, A
CC CONDITION CHARACTERIZED BY CIRCLING, HEAD-TOSSING, DEAFNESS AND
CC HYPERACTIVITY.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U09739; AAB00194.1; .
DR HSP; P08799; ILVK.
DR MGD; MGI:104785; MYO6.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 4.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; FALSE-NEG.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Disease mutation; Deafness.
FT DOMAIN 1 762 MYOSIN HEAD-LIKE.
FT DOMAIN 817 837 IQ.
FT DOMAIN 849 1014 COILED COIL (POTENTIAL).
FT NP_BIND 151 158 ATP (POTENTIAL).
FT DOMAIN 668 675 ACTIN-BINDING (POTENTIAL).
FT VARIANT 766 1265 MISSING (IN SNELL'S WALTZER).
FT SEQUENCE 1265 AA; 146408 MW; 4F5IABC72463148C CRC64;
SQ

```

```

Query Match 48.2%; Score 40; DB 1; Length 1265;
Best Local Similarity 56.2%; Pred. No. 1.4e+02;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 2 SAWQLAKQKQAEKL 17
DB 914 SALQKKQKEEAERL 929

RESULT 13
GSPH_ERMCH
ID GSPH_ERMCH STANDARD; PRT; 181 AA.
AC P24667;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE General secretion pathway protein H precursor (Pectic enzymes
DE secretion protein out).
GN OUTH.
OS Erwinia chrysanthemi.

```

```

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EC16;
RX MEDLINE=93054355; PubMed=1429461;
RA Lindeberg M., Collmer A.;
RT "Analysis of eight out genes in a cluster required for pectic enzyme
RT secretion by Erwinia chrysanthemi: sequence comparison with secretion
RT genes from other Gram-negative bacteria.";
RL J. Bacteriol. 174:7385-7397(1992).
RN [2]
RP SEQUENCE OF 85-181 FROM N.A.
RC STRAIN=EC16;
RX MEDLINE=91126059; PubMed=1992458;
RA He S.Y., Lindeberg M., Chatterjee A.K., Collmer A.;
RT "Cloned Erwinia chrysanthemi out genes enable Escherichia coli to
RT selectively secrete a diverse family of heterologous proteins to its
RT milieu.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:1079-1083(1991).
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF THE
CC MULTIPLE PECTIC ENZYMES.
CC -!- SIMILARITY: BELONGS TO THE PULH/OUTH/XPSH/EXEH/XCPU FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L02214; AAA24835.1; .
DR EMBL; M37886; AAA24826.1; .
DR PIR; A37874; A37874.
DR InterPro; IPR002416; Bac_GSPH.
DR InterPro; IPR001120; Prok_N.methyltn.
DR PRINTS; PR00885; BCTERIALGSPH.
DR PROSITE; PS00409; PROKAR_NTER_METHYL; 1.
KW Transport; Methylation.
FT PROPEP 1 5 BY SIMILARITY.
FT CHAIN 6 181 GENERAL SECRETION PATHWAY PROTEIN H.
FT MOD_RES 6 181 METHYLATION (BY SIMILARITY).
FT SEQUENCE 181 AA; 20288 MW; 7E72A931042B63B8 CRC64;
SQ

```

```

Query Match 47.0%; Score 39; DB 1; Length 181;
Best Local Similarity 63.6%; Pred. No. 31;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 2 SAWQLAKQKQAK 12
DB 37 SGWQIARFKAQ 47

RESULT 14
AANT_HDVSI
ID AANT_HDVSI STANDARD; PRT; 195 AA.
AC P25883;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Delta antigen.
OS Hepatitis delta virus (isolate Japanese S-1) (HDV), and
OS Hepatitis delta virus (isolate 7/18/83) (HDV).
OX Viruses; Deltavirus.
OX NCBI_TaxID=10427, 10421;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISOLATE JAPANESE S-1;
RX MEDLINE=91012805; PubMed=2214027;

```

```

A Imazeki F., Omata M., Ohto M.;
T "Heterogeneity and evolution rates of delta virus RNA sequences.";
N J. Virol. 64:5594-5599(1990).
L [2]
P SEQUENCE FROM N.A.
C STRAIN=ISOLATE 7/18/83;
X MEDLINE=9202044; PubMed=1923832;
A Imazeki F., Omata M., Ohto M.;
T "Complete nucleotide sequence of hepatitis delta virus RNA in Japan.";
L Nucleic Acids Res. 19:5439-5449(1991).
C -1- SUBCELLULAR LOCATION: Nuclear.
C -1- PTM: PHOSPHORYLATED.
C -1- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
C HEPATITIS DELTA VIRAL INFECTIONS.
C -1- MISCELLANEOUS: DELTA ANTIGEN BINDS SPECIFICALLY TO HDV RNA.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@sib-sib.ch).
C
C
R EMBL; D90193; BAA14217.1; .
R EMBL; X60193; CAA42749.1; .
R PIR; B36409; SAVLDS.
R PIR; S18678; S18678.
R InterPro: IPR002506; HDV_ag.
R Pfam; PF01517; HDV_ag; 1.
R ProDom; PD002887; HDV_ag; 1.
W Antigen; RNA-binding; Nuclear protein; Phosphorylation.
Q SEQUENCE 195 AA: 21928 MW: 40000 E032C20A CRC64;

```

```

Query Match 47.0%; Score 39; DB 1; Length 195;
Best Local Similarity 53.8%; Pred. No. 33;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

Y 4 WOLAKOKAQAEK 16
I ::::|:|:|
b 21 WITARKKAELEK 33

```

RESULT 15

```

ANT_HDVS2
D AANT_HDVS2 STANDARD; PRT; 195 AA.
C P25884;
T 01-MAY-1992 (Rel. 22, Created)
T 01-MAY-1992 (Rel. 22, last sequence update)
T 01-MAY-1992 (Rel. 22, last annotation update)
E Delta antigen.
S Hepatitis delta virus (isolate Japanese S-2) (HDV).
C Viruses; Deltavirus.
C NCBI_TaxID=10428;
N [1]
P SEQUENCE FROM N.A.
X MEDLINE=91012805; PubMed=2214027;
A Imazeki F., Omata M., Ohto M.;
T "Heterogeneity and evolution rates of delta virus RNA sequences.";
L J. Virol. 64:5594-5599(1990).
C -1- SUBCELLULAR LOCATION: Nuclear.
C -1- PTM: PHOSPHORYLATED.
C -1- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
C HEPATITIS DELTA VIRAL INFECTIONS.
C -1- MISCELLANEOUS: DELTA ANTIGEN BINDS SPECIFICALLY TO HDV RNA.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@sib-sib.ch).
C

```

```

CC or send an email to license@sib-sib.ch.
CC
DR EMBL; D90193; BAA14217.1; .
DR InterPro: IPR002506; HDV_ag.
DR Pfam; PF01517; HDV_ag; 1.
DR ProDom; PD002887; HDV_ag; 1.
KW Antigen; RNA-binding; Nuclear protein; Phosphorylation.
SQ SEQUENCE 195 AA: 21912 MW: 40000 E032C20A CRC64;

Query Match 47.0%; Score 39; DB 1; Length 195;
Best Local Similarity 53.8%; Pred. No. 33;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 WOLAKOKAQAEK 16
I ::::|:|:|
Db 21 WITARKKAELEK 33

Search completed: August 6, 2002, 17:07:35
Job time: 877 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

3M protein - protein search, using sw model

run on: August 6, 2002, 17:09:29 ; Search time 111.35 Seconds

(without alignments)

26.411 Million cell updates/sec

Title: US-10-020-139-2_COPY_68_84

Perfect score: 83

Sequence: 1 SSAWOLAKQKQAEKEL 17

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched:

562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	83	100.0	249	4 Q9BQQ0	Q9bqq0 homo sapien
2	83	100.0	249	4 Q96DR5	Q96dr5 homo sapien
3	46	55.4	858	16 Q9ZEA9	Q9zea9 rickettsia
4	44	53.0	1043	10 Q94BT0	Q94bt0 arabidopsis
5	44	53.0	1220	4 Q9UET5	Q9uet5 homo sapien
6	44	53.0	1220	4 Q9UNK1	Q9unk1 homo sapien
7	44	53.0	1721	4 Q95216	Q95216 homo sapien
8	44	53.0	1721	4 Q9UNK2	Q9unk2 homo sapien
9	43.5	52.4	475	13 Q9PUQ5	Q9puq5 opsanus bet
10	43	51.8	121	11 Q9D0Y8	Q9doy8 mus musculus
11	43	51.8	685	16 Q9JWP0	Q9jwp0 neisseria m
12	42	50.6	125	9 Q9T194	Q9t194 bacterioph
13	42	50.6	350	2 Q60253	Q60253 escherichia
14	42	50.6	731	5 Q9GRG0	Q9grg0 tetrahymena
15	42	50.6	736	10 Q23362	Q23362 arabidopsis
16	42	50.6	928	5 Q9VTN6	Q9vtn6 drosophila

17	42	50.6	1165	4 Q92619	Q92619 homo sapien
18	42	50.6	3115	5 Q95VT2	Q95vt2 leishmania
19	41	49.4	177	10 Q94IR1	Q94ir1 zea mays (m
20	41	49.4	185	16 Q67526	Q67526 aquifex aeo
21	41	49.4	271	4 Q9C031	Q9c031 homo sapien
22	41	49.4	302	4 Q9H1B0	Q9h1b0 homo sapien
23	41	49.4	305	4 Q9H1B1	Q9h1b1 homo sapien
24	41	49.4	326	4 Q9C032	Q9c032 homo sapien
25	41	49.4	347	4 Q9C033	Q9c033 homo sapien
26	41	49.4	368	4 Q9H1B3	Q9h1b3 homo sapien
27	41	49.4	397	4 Q9H1B4	Q9h1b4 homo sapien
28	41	49.4	400	4 Q9C034	Q9c034 homo sapien
29	41	49.4	409	4 Q9NS11	Q9ns11 homo sapien
30	41	49.4	493	4 Q9C035	Q9c035 homo sapien
31	41	49.4	493	4 Q96SR5	Q96sr5 homo sapien
32	41	49.4	515	4 Q9NX66	Q9nx66 homo sapien
33	41	49.4	560	4 Q9BXU4	Q9bxu4 homo sapien
34	41	49.4	614	10 Q9LI68	Q9li68 arabidopsis
35	41	49.4	625	4 Q9G2Y0	Q9gy20 homo sapien
36	41	49.4	857	16 Q9ZJK8	Q9zjk8 rickettsia
37	41	49.4	1148	16 Q9HZK3	Q9hzk3 pseudomonas
38	41	49.4	1992	3 Q9P6T1	Q9p6t1 neurospora
39	40	48.2	125	16 Q92B53	Q92b53 listeria in
40	40	48.2	150	15 Q66724	Q66724 equine infe
41	40	48.2	158	10 Q9SMP3	Q9sm3 arabidopsis
42	40	48.2	185	2 Q87122	Q87122 actinobacill
43	40	48.2	186	2 Q9F280	Q9f280 actinobacill
44	40	48.2	186	2 Q06524	Q06524 haemophilus
45	40	48.2	190	16 Q9K8H3	Q9k8h3 bacillus ha

ALIGNMENTS

RESULT 1
Q9BQQ0 PRELIMINARY: PRT; 249 AA.
AC Q9BQQ0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BA49G10.1 (SIMILAR TO BOVINE SALIVARY PROTEIN BSP30).
GN BA49G10.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tracey A.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL121901; CAC03546.1; -.
SQ SEQUENCE 249 AA; 27011 MW; E64E0794A1B4DB7D CRC64;

Query Match 100.0%; Score 83; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSAWOLAKQKQAEKEL 17

Db 68 SSAWOLAKQKQAEKEL 84

RESULT 2

Q96DR5 PRELIMINARY: PRT; 249 AA.

AC Q96DR5;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DE PAROTID SECRETORY PROTEIN.

OS Homo sapiens (Human).

SULT 4
4BT0

```

RC TISSUE=BRAIN;
RX MEDLINE=99017974; PubMed=9799604;
RA Guipponi M., Scott H.S., Chen H., Schebesta A., Rossier C.,
RA Antonarakis S.E.;
RT "Two isoforms of a human intersectin (ITSN) protein are produced by
RT brain-specific alternative splicing in a stop codon.";
RL Genomics 53:369-376(1998).
CC -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
DR EMBL: AF064244; AAC78611.1; -.
DR HSSP: P29355; ISEM.
DR InterPro: IPR000008; C2.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000261; EPS15_repeat.
DR InterPro: IPR000108; Neu_cyt_fact_2.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00168; C2; 1.
DR Pfam: PF00036; ehand; 3.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhoGEF; 1.
DR Pfam: PF00018; SH3; 5.
DR PRINTS: PR00499; P67PHOX.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART: SM00239; C2; 1.
DR SMART: SM00054; Eph; 2.
DR SMART: SM00027; EH; 2.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhoGEF; 1.
DR SMART: SM00326; SH3; 5.
DR PROSITE: PS50004; C2_DOMAIN_2; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS50002; SH3; 5.
DR Calcium-binding.
KW SEQUENCE 1721 AA; 195559 MW; 851A5CFB2BC4EBFC CRC64;
SQ

Query Match 53.0%; Score 44; DB 4; Length 1721;
Best Local Similarity 69.2%; Pred. No. 2e+02; Mismatches 2; Indels 0; Gaps 0;
Matches 9; Conservative 2;

QY 5 QLAKQKQAEAKL 17
Db 618 QLQKQKMEAEAL 630
|||:||||:

RESULT 8
Q9UNK2 PRELIMINARY; PRT; 1721 AA.
AC Q9UNK2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE INTERSECTIN LONG ISOFORM.
GN ITSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pucharcos C., Fuentes J.J., Casas C., de la Luna S., Alcantara S.,
RA Arbones M.L., Soriano E., Estivill X., Pritchard M.;
RT "Alu-splice cloning of human intersectin (ITSN), a putative
RT multivalent binding protein expressed in proliferating and
RT differentiating neurons and overexpressed in Down syndrome.";
RL Eur. J. Hum. Genet. 0:0-0(1999).
CC -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
DR EMBL: AF114487; AAD29952.1; -.
DR HSSP: P29355; ISEM.
DR InterPro: IPR000008; C2.
DR InterPro: IPR002048; EF-hand.
KW SEQUENCE FROM N.A.
SQ

Query Match 53.0%; Score 44; DB 4; Length 1220;
Best Local Similarity 69.2%; Pred. No. 1.4e+02;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 5 QLAKQKQAEAKL 17
Db 618 QLQKQKMEAEAL 630
|||:||||:

RESULT 7
O95216 PRELIMINARY; PRT; 1721 AA.
AC O95216;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE INTERSECTIN LONG FORM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
SQ

```

R InterPro: IPR000261; EPS15_repeat.
 R InterPro: IPR000108; Neu_cyt_fact_2.
 R InterPro: IPR001849; PH.
 R InterPro: IPR000219; RhoGEF.
 R InterPro: IPR001452; SH3.
 R Pfam: PF00168; C2; 1.
 R Pfam: PF00036; efnand; 3.
 R Pfam: PF00169; PH; 1.
 R Pfam: PF00621; RhoGEF; 1.
 R Pfam: PF00018; SH3; 5.
 R PRINTS: PR00499; P67PHOX.
 R PRINTS: PR00452; SH3DOMAIN.
 R SMART: SM00239; C2; 1.
 R SMART: SM00054; Efn; 2.
 R SMART: SM00027; EH; 2.
 R SMART: SM00233; PH; 1.
 R SMART: SM00325; RhoGEF; 1.
 R SMART: SM00326; SH3; 5.
 R PROSITE: PS00004; C2_DOMAIN_2; 1.
 R PROSITE: PS00018; EF_HAND; UNKNOWN_2.
 R PROSITE: PS00003; PH_DOMAIN; 1.
 R PROSITE: PS00002; SH3; 5.
 W Calcium-binding.
 Q SEQUENCE 1721 AA; 195421 MW; FC4DE644D8BEA2BE CRC64;

Query Match 53.0%; Score 44; DB 4; Length 1721;
 Best Local Similarity 69.2%; Pred. No. 26+02;
 Matches 9; Conservative 2; Mismatches 0; Gaps 0;

y 5 QLAKQKQAEAEKL 17
 |||||
 b 618 QLQKQKSMEEARL 630

RESULT 9
 QP005 PRELIMINARY; PRT: 475 AA.

C QP005
 T 01-MAY-2000 (TEMBLrel. 13, Created)
 T 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 T 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 E UREA TRANSPORTER.
 S Opsonus beta.
 C Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 C Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 C Acanthomorpha; Paracanthopterygii; Batrachoididae; Opsanus.
 X NCBI_TaxID=95145;
 N [1]
 P SEQUENCE FROM N.A.
 C TISSUE-GILL;
 X MEDLINE=20347328; PubMed=10887074;
 A Walsh P.J., Heitz M.J., Campbell C.E., Cooper G.J., Medina M.,
 A Wang Y.S., Goss G.G., Vincek V., Wood C.M., Smith C.P.;
 T "Molecular characterization of a urea transporter in the gill of the
 T gulf toadfish."
 L J. Exp. Biol. 203:2357-2364(2000).
 L ENBL: AF165893; AAD53268.2;
 R InterPro: IPR001117; Cu-oxidase.
 R PROSITE: PS00079; MULTICOPPER_OXIDASE1; UNKNOWN_1.
 Q SEQUENCE 475 AA; 52838 MW; 700E2B180445198C CRC64;

Query Match 52.4%; Score 43.5; DB 13; Length 475;
 Best Local Similarity 66.7%; Pred. No. 61;
 Matches 10; Conservative 2; Mismatches 1; Gaps 1;

y 4 WLAKQKQAE-AEKL 17
 |||||
 b 362 WLKLKQKQKQAEADKL 396

RESULT 10

Q9D0Y8 PRELIMINARY; PRT: 121 AA.
 AC Q9D0Y8
 DT 01-JUN-2001 (TEMBLrel. 17, Created)
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE 1110047B07RIK PROTEIN.
 GS MRPL52 OR 1110047B07RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK004194; BAB23216.1;
 DR MGD: MGI:1916086; Mrpl52.
 SQ SEQUENCE 121 AA; 13658 MW; C024F28F8A5B016A CRC64;

Query Match 51.8%; Score 43; DB 11; Length 121;
 Best Local Similarity 57.1%; Pred. No. 17;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 AWLAKQKQAEAEK 16
 ||:|:|:|:|
 DB 89 ANKLQKQKQAEK 102

RESULT 11
 Q9JWP0 PRELIMINARY; PRT: 685 AA.
 AC Q9JWP0
 DT 01-OCT-2000 (TEMBLrel. 15, Created)
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10).
 GS METG OR NMA0275.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Batrell B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria

RT meningitidis Z2491.1;
 RL Nature 404:502-506(2000).
 DR EMBL: AL162752; CAB83583.1;
 DR HSP: P00959; IMEA
 DR InterPro: IPR001412; LRNA-synt_1.
 DR InterPro: IPR002304; LRNA-synt_met.
 DR InterPro: IPR002547; tRNA_bind.
 DR Pfam: PF01588; LRNA_bind; 1.
 DR PRINTS: PRO1041; TRNASYNTHET.
 DR PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.
 KW Aminoacyl-tRNA synthetase; Ligase; Complete proteome.
 SQ SEQUENCE 685 AA; 76984 MW; 07FD45915ED3BEF3 CRC64;

Query Match 51.8%; Score 43; MB 16; Length 685;
 Best Local Similarity 63.6%; Pred. No. 1;le-02;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 4 WOLAKOKAQEA 14
 I:||||:|I
 Db 468 WELAKQEQQA 478

RESULT 12
 Q9TI94
 ID Q9TI94 PRELIMINARY; PRT; 125 AA.
 AC Q9TI94
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE GP30.
 OS Bacteriophage A118.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 OX NCBI_TaxID=40521;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Loessner M.J.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96020653; PubMed=8577256;
 RA Loessner M.J., Wendlinger G., Scherer S.;
 RT "heterogeneous endolysins in *Listeria monocytogenes* bacteriophages: a
 RT new class of enzymes and evidence for conserved holin genes within the
 RT siphoviral lysis cassettes";
 RL Mol. Microbiol. 16:1231-1241(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20117992; PubMed=10652093;
 RA Loessner M.J., Innan R.B., Lauer P., Calendar K.;
 RT "Complete nucleotide sequence, molecular analysis and genome structure
 RT of bacteriophage A118 of *Listeria monocytogenes*: implications for
 RT phage evolution";
 RL Mol. Microbiol. 35:324-340(2000).
 DR EMBL: AJ242593; CAB53816.1;
 SQ SEQUENCE 125 AA; 14707 MW; AF95EPPFAE323ADF0 CRC64;

Query Match 50.6%; Score 42; DB 9; Length 125;
 Best Local Similarity 46.7%; Pred. No. 25;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SAMOLAKOKAQEA 16
 I:||||:|I
 Db 43 AAWSKAEVKEV 57

RESULT 13
 Q60253
 ID Q60253 PRELIMINARY; PRT; 350 AA.
 AC Q60253; Q47436; Q47434; Q47435;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 5K PA GENE ENCODING PENICILLIN ACYLASE, SEGMENT 1 PRECURSOR
 OS (FRAGMENT).
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85236066; PubMed=2989404;
 RA Bruns W., Hoppe J., Tsai H., Bruening H.J., Maywald F., Collins J.,
 RA Mayer H.;
 RT "Structure of the penicillin acylase gene from *Escherichia coli*: A
 RT periplasmic enzyme that undergoes multiple proteolytic processing";
 RL J. Mol. Appl. Genet. 3:36-44(1985).
 DR EMBL: M11672; AAA24258.1;
 DR HSP: P06875; IAJQ.
 DR InterPro: IPR002692; Penicill_amidase.
 DR Pfam: PF01804; Penicill_amidase; 1.
 KW Signal.
 FT SIGNAL. 1 26 POTENTIAL.
 FT CHAIN 27 >227 POTENTIAL.
 FT CHAIN 27 >235 POTENTIAL.
 FT CHAIN 290 >350 POTENTIAL.
 FT NON_TER 350 350
 SQ SEQUENCE 350 AA; 38903 MW; 1D97970699A751DC CRC64;

Query Match 50.6%; Score 42; DB 2; Length 350;
 Best Local Similarity 41.2%; Pred. No. 77;
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SSANOLAKOKAQEA 17
 I:||||:|I
 Db 290 SNMWVUKSQAQDAKAI 306

RESULT 14
 Q9GRG0
 ID Q9GRG0 PRELIMINARY; PRT; 731 AA.
 AC Q9GRG0
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE TETRIN B PROTEIN.
 OS Tetrahymena thermophila.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
 OC Tetrahymenina; Tetrahymena.
 OX NCBI_TaxID=5911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SB281;
 RX MEDLINE=20419140; PubMed=10965956;
 RA Brimmer A., Weber K.;
 RT "The cDNA sequences of three tetris, the structural proteins of the
 RT Tetrahymena oral filaments, show that they are novel cytoskeletal
 RT proteins";
 RL Protist 151:171-180(2000).
 DR EMBL: AJ276471; CAB07819.1;
 SQ SEQUENCE 731 AA; 87196 MW; A2FC64972BF06BE0 CRC64;

Query Match 50.6%; Score 42; DB 5; Length 731;
 Best Local Similarity 69.2%; Pred. No. 1.7e+02;
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 5 QIAKOKAQEA 17
 I:||||:|I
 Db 386 QIWKEKAAEA 398

RESULT 15
 Q2362

```

D O23362 PRELIMINARY: PRT: 736 AA.
C O23362:
Y 01-JAN-1998 (TrEMBLrel. 05, Created)
Y 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
Y 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
Y HYPOTHEtical 80.8 KDA PROTEIN.
N AT4G15080.
S Arabidopsis thaliana (Mouse-ear cross).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
C eurosids II; Brassicales; Brassicaceae; Arabidopsids.
X NCBI_TaxID=3702;
X [1]
X [2]
X SEQUENCE FROM N.A.
X MEDLINE=98121113; PubMed=9461215;
X Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
X Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
X Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
X Wedler E., Wambutt R., Weitzenecker T., Pohl T.M., Terry N.,
X Gielen J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
X Auborg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
X Entian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
X Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
X Vouklatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
X Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
X Palme K., Benes V., Rechman S., Ansoorge W., Cooke R., Berger C.,
X Delseny M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
X Schueller C., Chalwatzis N.;
X "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
X Arabidopsis thaliana.";
X Nature 391:485-488(1998).
X [2]
X SEQUENCE FROM N.A.
X EU Arabidopsis sequencing project:
X Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
X EMBL: Z97337; CAB10287.1; -.
X EMBL: AL161540; CAB78550.1; -.
X InterPro: IPR001594; Znf-DHHC.
X Pfam: PF01529; zf-DHHC; 1.
X ProDom: PD003041; Znf-DHHC; 1.
X Hypothetical protein.
X SEQUENCE 736 AA; 80802 MW; 92D443CDFEE20567 CRC64;

```

```

Query Match 50.6%; Score 42; DB 10; Length 736;
Best Local Similarity 53.3%; Pred. NO. 1.7e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

ZY 2 SAWOLAKOQAEK 16
   |||:|:|:|:|:|
ZY 422 SAWKLKLSNEATR 436

```

```

Search completed: August 6, 2002, 17:09:30
Job time: 932 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: August 6, 2002, 16:52:52 ; Search time 138.55 Seconds

(without alignments)
13.629 Million cell updates/sec

Title: US-10-020-139-2_COPY_68_84

Perfect score: 83

Sequence: 1 SSAWQLAKKQAEKEL 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1980.DAT.*
- 2: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1981.DAT.*
- 3: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1982.DAT.*
- 4: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1983.DAT.*
- 5: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1984.DAT.*
- 6: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1985.DAT.*
- 7: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1986.DAT.*
- 8: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1987.DAT.*
- 9: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1988.DAT.*
- 10: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1989.DAT.*
- 11: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1990.DAT.*
- 12: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1991.DAT.*
- 13: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1992.DAT.*
- 14: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1993.DAT.*
- 15: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1994.DAT.*
- 16: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1995.DAT.*
- 17: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1996.DAT.*
- 18: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1997.DAT.*
- 19: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1998.DAT.*
- 20: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1999.DAT.*
- 21: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA2000.DAT.*
- 22: /SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	83	100.0	249	19	AAW69221 Human parotid secr
2	83	100.0	249	19	AAW60682 Human parotid secr
3	83	100.0	249	21	AA324069 Human PRO1025 prot
4	83	100.0	249	21	AA325765 Human secreted pro
5	83	100.0	249	22	AA375351 Human secreted pro
6	83	100.0	260	22	AA25745 Human protein sequ
7	46	55.4	85	22	AAW90448 Human immune/haema
8	46	55.4	123	22	AA504203 Human gene 5 encod
9	46	55.4	123	22	AA504221 Human gene 5 encod
10	44	53.0	595	22	AA394691 Human protein sequ
11	44	53.0	641	20	AAV32158 Human SH3D1A prote

12	44	53.0	642	22	AAW93229 Human polypeptide,
13	44	53.0	648	20	AAV32157 Human SH3D1A prote
14	44	53.0	932	22	ABG20575 Novel human diagno
15	44	53.0	1035	22	AA431519 Human polypeptide
16	44	53.0	1144	20	AAV32154 Human SH3D1A prote
17	44	53.0	1215	20	AAV32156 Human SH3D1A prote
18	44	53.0	1220	20	AAV32155 Human SH3D1A prote
19	42	50.6	34	21	AAV50199 Streptococcus pyog
20	42	50.6	557	19	AAW57840 Wild type penicilli
21	42	50.6	557	19	AAW57843 V56T mutant penici
22	42	50.6	557	19	AAW57844 F24A mutant penici
23	42	50.6	557	19	AAW57845 F24V mutant penici
24	42	50.6	557	19	AAW57846 F24L mutant penici
25	42	50.6	846	7	AAV61009 Sequence encoded b
26	42	50.6	846	22	AAV37817 A. faecalis-E. col
27	42	50.6	928	22	ABV61966 Drosophila melanog
28	41	49.4	493	22	ABV5243 Human protein sequ
29	41	49.4	560	22	AAU07884 Polypeptide sequen
30	41	49.4	1045	22	AAU35024 Enterococcus faeca
31	41	49.4	1148	22	AAU36302 Pseudomonas aerugi
32	40	48.2	52	21	AAV21205 Arabidopsis thalia
33	40	48.2	54	21	AAV21204 Arabidopsis thalia
34	40	48.2	113	21	AAV21204 Arabidopsis thalia
35	40	48.2	158	21	AAV21204 Arabidopsis thalia
36	40	48.2	230	22	ABV58344 Drosophila melanog
37	40	48.2	232	22	ABV58317 Drosophila melanog
38	40	48.2	396	19	AAW30682 S. pneumoniae prot
39	40	48.2	419	22	ABV70070 Drosophila melanog
40	40	48.2	671	21	AAV74342 Neisseria gonorrhe
41	39.5	47.6	40	18	AAW19086 Trypanosoma cruzi
42	39.5	47.6	40	20	AAV23282 Repeat sequence of
43	39.5	47.6	40	21	AAV26481 Synthetic peptide
44	39.5	47.6	186	18	AAW26536 Trypanosoma cruzi
45	39.5	47.6	186	20	AAV23298 Trypanosoma cruzi a

ALIGNMENTS

RESULT 1	AAW69221	AAW69221 standard; Protein; 249 AA.
ID	AAW69221	standard; Protein; 249 AA.
XX	AAW69221	
AC	AAW69221	
XX	16-OCT-1998	(first entry)
DT	16-OCT-1998	(first entry)
XX	Human parotid	secretory protein.
DE	Human parotid	secretory protein.
XX	Parotid secretory	protein; hsp; digestive disorder; endocrine disorder;
KW	non-immune	defensive disorder; immune system disorder; cancer; human;
KW	therapy:	diagnosis.
XX	Homo sapiens.	
OS	Homo sapiens.	
XX	Key	Location/Qualifiers
FT	Peptide	1..18
FT	Protein	/note= "signal peptide"
FT	Protein	19..249
FT	Protein	/note= "mature hPSP"
XX	WO9828420-A1.	
PN	WO9828420-A1.	
XX	02-JUL-1998.	
PD	02-JUL-1998.	
XX	18-DEC-1997;	97WO-US23522.
PF	18-DEC-1997;	97WO-US23522.
XX	23-DEC-1996;	96US-0034429.
PR	23-DEC-1996;	96US-0034429.
XX	(HUMA-)	HUMAN GENOME SCI INC.
PA	(HUMA-)	HUMAN GENOME SCI INC.
XX	Duan R.	Ruben SM.
PI	Duan R.	Ruben SM.
XX		

WPI: 1998-377651/32.
 N-PSDB; AAV44759.

New nucleic acid encoding human parotid secretory protein or its fragments - useful for diagnosis and treatment of, e.g. digestive and endocrine disorders and for drug screening

Claim 16; Fig 1; 94pp; English.

This sequence is the human parotid secretory protein (hPSP) of the invention. The hPSP DNA is useful for chromosome identification and isolation of the corresponding genomic DNA. The DNA and protein can be used to detect abnormal levels of hPSP (in standard blotting, amplification or immuno assays), particularly for diagnosis of digestive, non-immune defensive, endocrine or immune system disorders. A particular application is diagnosis of cancers of the salivary gland, thymus and pancreas which are associated with high levels of hPSP. The protein is also useful as antifungal, antibacterial, antiparasitic and antiviral agents and may be expressed in vivo from the DNA. The protein, or cells expressing it, are used in screening tests to identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and antisense nucleic acids, which are potentially useful for treating conditions associated with excessive hPSP production. Cells containing the DNA are used to express the recombinant protein and this can be used to raise Ab, useful for diagnosis, therapy, for affinity purification and to identify hPSP-binding proteins.

Sequence 249 AA;

Query Match 100.0%; Score 83; DB 19; Length 249;
 Best Local Similarity 100.0%; Pred. No. 5.9e-05;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 SSAWQLAKKQAEAEKL 17
 |||||
 D 68 ssawqlakqkaeakl 84

RESULT 2
 AW60682
 J AAW60682 standard; Protein; 249 AA.
 X AAW60682;
 X
 F 18-SEP-1998 (first entry)
 X
 X Human parotid secretory protein (HPSP).
 E
 X Parotid secretory protein; human; cancer; autoimmune disease;
 X secretory tissue; gastrointestinal tissue; HPSP; Sjogren's syndrome;
 X Graves disease; thyroiditis; insulin-dependent diabetes; pancreatitis;
 X ulcerative colitis; Crohn's disease; atrophic gastritis.
 X
 S Homo sapiens.
 X
 X WO9821329-A1.
 X
 D 22-MAY-1998.
 X
 X 07-NOV-1997; 97WO-US20651.
 F
 X 14-NOV-1996; 96US-0749288.
 R
 X (INCY-) INCYTE PHARM INC.
 A
 X Bandman O, Goli SK;
 I
 X WPI: 1998-297933/26.
 X
 X N-PSDB; AAV37699.
 X
 X New parotid secretory protein - useful for, e.g. treatment of cancer and auto-immune disease, particularly of secretory or

PT gastrointestinal tissues
 XX
 PS Claim 1; Fig 1A-C; 65pp; English.
 XX
 CC This represents a human parotid secretory protein (HPSP). Antagonists that bind specifically to, and modulate activity of HPSP are used to treat cancer and autoimmune diseases particularly of secretory or gastrointestinal tissue, e.g. cancer of salivary gland, thyroid, prostate, breast, gastrointestinal tract or pancreas, Sjogren's syndrome, Graves disease, thyroiditis, insulin-dependent diabetes, pancreatitis, ulcerative colitis, Crohn's disease and atrophic gastritis. Host cells containing expression vectors comprising the HPSP nucleic acid are used to produce recombinant HPSP which is used to generate antibodies and to screen for its antagonists. Antibodies are useful directly as antagonists, to transport drugs to HPSP-expressing cells, to detect cells that express HPSP, to monitor patients being treated with HPSP, and for purification of HPSP from natural sources. Expression of HPSP may indicate cell proliferation. HPSP nucleic acid or its fragments are used to detect HPSP-encoding sequences (optionally after amplification by PCR) by hybridisation, particularly for diagnosis and monitoring of disease, but also for mapping the chromosomal sequence.

XX Sequence 249 AA;

Query Match 100.0%; Score 83; DB 19; Length 249;
 Best Local Similarity 100.0%; Pred. No. 5.9e-05;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSAWQLAKKQAEAEKL 17
 |||||
 Db 68 ssawqlakqkaeakl 84

RESULT 3
 AAB24069
 ID AAB24069 standard; Protein; 249 AA.
 XX
 AC AAB24069;
 XX
 X 29-JAN-2001 (first entry)
 XX
 DE Human PRO1025 protein sequence SEQ ID NO:38.
 XX
 X Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth; proliferation; tumorigenesis; identification; cancer; cytostatic; nontropic; neuroprotective; antiinflammatory; immunosuppressive; immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy; neuronal disorder; glial disorder; astrocytal disorder; angiogenic; hypothalamic disorder; glandular disorder; macrophagal disorder; epithelial disorder; stromal disorder; blastocoele disorder; inflammatory disorder; immunologic disorder.
 XX
 OS Homo sapiens.
 XX
 X WO2000053755-A2.
 XX
 PD 14-SEP-2000.
 XX
 X 06-JAN-2000; 2000WO-US00376.
 PF
 XX 08-MAR-1999; 99WO-US05028.
 PR 02-JUN-1999; 99WO-US12252.
 PR 23-JUN-1999; 99US-0141037.
 PR 07-JUL-1999; 99US-0143048.
 PR 26-JUL-1999; 99US-0145698.
 PR 30-NOV-1999; 99WO-US28313.
 PR 20-DEC-1999; 99WO-US30911.
 PR 05-JAN-2000; 2000WO-US00219.
 XX
 X (GETH) GENENTECH INC.
 PA
 XX Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;
 PI

2I Watanabe CX, Wood WI;
 XX WPI: 2000-572270/53.
 JR N-PSDB; AAC58379.
 XX Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
 PT treatment, diagnosis and prevention of cancer.
 XX
 PS Claim 61; Fig 26: 286pp; English.
 XX
 CC The present invention describes an isolated antibody that binds to
 CC one of the human PRO proteins designated PRO12, PRO290, PRO341, PRO535,
 CC PRO619, PRO717, PRO809, PRO848, PRO943, PRO1005, PRO1009,
 CC PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
 CC PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
 CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
 CC growth. The PRO polypeptides and nucleotides are useful in the
 CC treatment, diagnosis and prevention of cancer. The antibodies and other
 CC anti-tumour compounds may be used to treat various conditions, including
 CC those characterised by overexpression and/or activation of the amplified
 CC PRO genes. Exemplary conditions or disorders to be treated with such
 CC antibodies and other compounds include benign or malignant tumours
 CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
 CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
 CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
 CC leukaemias and lymphoid malignancies, other disorders such as neuronal,
 CC gliol, astrocytal, hypothalamic and other glandular, macrophagal,
 CC epithelial, stromal and blastocoeic disorders, and inflammatory,
 CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
 CC primers and hybridisation probes used in the isolation of the human PRO
 CC sequences. AAC58367 to AAC58396 and AAC24057 to AAC24089 represent human
 CC PRO polynucleotide and protein sequences given in the exemplification of
 CC the present invention.
 XX
 XX Sequence 249 AA;
 SQ

Query Match 100.0%; Score 83; DB 21; Length 249;
 Best Local Similarity 100.0%; Pred. No. 5.9e-05;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 0;
 0;

QY 1 SSAWOLAKKQAEKEL 17
 |||||
 DB 68 ssawqlakgkqaeakl 84

RESULT 4
 AAB25765
 ID AAB25765 standard; Protein: 249 AA.
 AC AAB25765;
 XX
 XX 28-NOV-2000 (first entry)
 DT Human secreted protein SEQ ID #77.
 DE
 DE Human; secreted protein; forensic procedure; gene therapy;
 KW chromosome mapping; cancer; autoimmune disease; cardiovascular disorder;
 KW cystic fibrosis; hypothyroidism; immunological disorder; amyloidosis;
 KW brain disorder; skeletal muscle disorder; eye disorder; obesity;
 KW mitochondrialopathy; diabetes; atherosclerosis; Alzheimer's disease;
 KW neurodegenerative disorder; graft rejection; dementia; hyperlipidaemia;
 KW septic shock; impotence.
 XX
 OS Homo sapiens.
 XX
 XX WO200037491-A2.
 PN
 XX 29-JUN-2000.
 PD
 XX 20-DEC-1999; 99WO-1B02058.
 PF
 XX 22-DEC-1998; 98US-0113686.
 PR

PR 25-JUN-1999; 99US-0141032.
 XX (GEST) GENSET.
 PA Bougueleret L, Dumas J, Duclert A;
 PI WPI: 2000-442637/38.
 XX N-PSDB; AAA87727.
 DR
 DR Polynucleotides and polypeptides encoding proteins with signal
 PT mapping procedures -
 PT
 PS Claim 9; Figure 10; 306pp; English.
 XX
 CC This sequence represents a human secreted protein amino acid sequence.
 CC The invention relates to sequences AAA87725-A87774 which encode human
 CC secreted proteins AAB25763-B25812. The proteins include signal peptides.
 CC Included in the invention are a host cell containing one of the cDNA
 CC sequences, and a purified antibody capable of binding to one of the
 CC secreted proteins. Also contained in the invention are methods for
 CC storing the sequence data on a computer system, and a method for
 CC identifying features of the cDNA sequences using a computer programme.
 CC The cDNAs are useful for expressing secreted proteins or fragments to
 CC obtain antibodies capable of specifically binding to the secreted
 CC proteins. The cDNAs may also be useful in diagnostic, forensic, gene
 CC therapy and chromosome mapping procedures and may be used to design
 CC expression vectors and secretion vectors. The proteins of the invention
 CC may be used to treat diseases including cancer, autoimmune diseases,
 CC cardiovascular disorders, cystic fibrosis, hypothyroidism, immunological
 CC disorders, amyloidosis, brain disorders, skeletal muscle disorders, eye
 CC disorders, obesity, mitochondrialocytopathies, diabetes, atherosclerosis,
 CC neurodegenerative disorders, graft rejection, Alzheimer's disease,
 CC dementia, hyperlipidaemia, septic shock and impotence.
 XX
 XX Sequence 249 AA;
 SQ

Query Match 100.0%; Score 83; DB 21; Length 249;
 Best Local Similarity 100.0%; Pred. No. 5.9e-05;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 0;
 0;

QY 1 SSAWOLAKKQAEKEL 17
 |||||
 DB 68 ssawqlakgkqaeakl 84

RESULT 5
 AAB75351
 ID AAB75351 standard; protein: 249 AA.
 XX
 AC AAB75351;
 XX
 XX 05-APR-2001 (first entry)
 DT Human secreted protein #10.
 DE
 DE Secreted protein; prevention; treatment; diagnosis; disease;
 KW infection.
 KW
 XX Homo sapiens.
 OS
 XX WO200100806-A2.
 PN
 XX 04-JAN-2001.
 PD
 XX 21-JUN-2000; 2000WO-IB00951.
 PF
 XX 25-JUN-1999; 99US-0141032.
 PR
 XX 21-DEC-1999; 99US-0469095.
 XX
 XX (GEST) GENSET.
 PA
 XX

1 Dumas Milne Edwards J, Bougueleret L, Jobert S;
2 WPI; 2001-071487/08.

3 49 Secreted proteins and the nucleic acids encoding them, useful in
4 gene therapy and for detecting similar sequences in samples -

5 Claim 10; Page 281; 307pp; English.

6 The present invention relates to 49 secreted proteins and the cDNAs
7 encoding them. The protein and nucleic acids may be used in the
8 prevention, treatment and diagnosis of diseases associated with
9 inappropriate protein expression.

10 Sequence 249 AA;

Query Match 100.0%; Score 83; DB 22; Length 249;

Best Local Similarity 100.0%; Pred. No. 5.9e-05;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 1 SSAWLAKKQKQAEAKL 17

2 68 ssawqlakkaqaekl 84

RESULT 6

AM25745

1 AAM25745 standard; Protein: 260 AA.

2 AAM25745;

3 16-OCT-2001 (first entry)

4 Human protein sequence SEQ ID NO:1260.

5 Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
6 antinflammatory; antirheumatic; antiarthritic; immunosuppressive;
7 antibacterial; endocrine; cardiant; central nervous system; virucide;
8 anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
9 antiaggregant; haemostatic; vulnery; antitumor; osteopathic; eczema;
10 dermatological; anti-allergic; antiasthmatic; antidiabetic; cytostatic;
11 neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
12 immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;
13 antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
14 cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
15 genetic disease; haematopoietic disorder; platelet disorder; asthma;
16 thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
17 allergic rhinitis; diabetes; multiple sclerosis; depression;
18 Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
19 neurological disorder.

20 Homo sapiens.

21 WO200153455-A2.

22 26-JUL-2001.

23 22-DEC-2000; 2000WO-US35017.

24 23-DEC-1999; 99US-0471275.

25 21-JAN-2000; 2000US-0488725.

26 25-APR-2000; 2000US-0552317.

27 (HYSE-) HYSEQ INC.

28 Tang YT, Liu C, Drmanac RT;

29 WPI; 2001-457603/49.

30 N-1-SDB; AAH99686.

31 Isolated human polynucleotides encoding polypeptides useful for the
32 treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX

PS

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Claim 20; Page 260; 1217pp; English.

AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
AAM25963. The proteins can have activities based on the tissues and
cells they are expressed in, such as: antinflammatory; antirheumatic;
antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
central nervous system; virucide; anti-HIV; fungicide; antitumor;
cardiovascular; antianaemic; antiaggregant; antidiabetic; vulnery;
antitumor; osteopathic; dermatological; anti-allergic; antisthmatic;
antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
antiparkinsonian; and immunostimulant. The proteins and polynucleotides
encoding them can be used in gene therapy, antitense therapy and vaccine
production. The proteins and polynucleotides are useful for screening for
agonists or antagonists of a protein and for the treatment and diagnosis
of disorders associated with the activity of a protein e.g. inflammation,
rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
infections, autoimmunity, genetic diseases, haematopoietic disorders,
anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
osteoporosis, severe combined immunodeficiency, eczema, allergic
rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
Alzheimer's disease, Parkinson's disease, neurodegenerative and
neurological disorders.

Sequence 260 AA;

Query Match 100.0%; Score 83; DB 22; Length 260;

Best Local Similarity 100.0%; Pred. No. 6.2e-05;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 1 SSAWLAKKQKQAEAKL 17

2 79 ssawqlakkaqaekl 95

RESULT 7

AAM90448

1 AAM90448 standard; Protein: 85 AA.

2 AAM90448;

3 07-NOV-2001 (first entry)

4 Human immune/haematopoietic antigen SEQ ID NO:18041.

5 Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
6 cytostatic; gene therapy; vaccine; metastasis.

7 Homo sapiens.

8 WO200157182-A2.

9 09-AUG-2001.

10 17-JAN-2001; 2001WO-US01354.

11 31-JAN-2000; 2000US-0179065.

12 04-FEB-2000; 2000US-0180628.

13 24-FEB-2000; 2000US-0184664.

14 02-MAR-2000; 2000US-0186350.

15 16-MAR-2000; 2000US-0189874.

16 17-MAR-2000; 2000US-0190076.

17 18-APR-2000; 2000US-0198123.

18 19-MAY-2000; 2000US-0205515.

19 07-JUN-2000; 2000US-0209467.

20 28-JUN-2000; 2000US-0214886.

21 30-JUN-2000; 2000US-0215135.

22 07-JUL-2000; 2000US-0216647.

23 07-JUL-2000; 2000US-0216680.

24 11-JUL-2000; 2000US-0217487.

25 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-02182290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231143.
PR 08-SEP-2000; 2000US-0231144.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-02344223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0245474.
PR 08-NOV-2000; 2000US-0245475.
PR 08-NOV-2000; 2000US-0245476.
PR 08-NOV-2000; 2000US-0245477.
PR 08-NOV-2000; 2000US-0245478.
PR 08-NOV-2000; 2000US-0245523.
PR 08-NOV-2000; 2000US-0245524.
PR 08-NOV-2000; 2000US-0245525.
PR 08-NOV-2000; 2000US-0245526.
PR 08-NOV-2000; 2000US-0245527.
PR 08-NOV-2000; 2000US-0245528.
PR 08-NOV-2000; 2000US-0245532.
PR 08-NOV-2000; 2000US-0245609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2000US-0254097.
PR 05-JAN-2001; 2000US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI: 2001-483426/52.

N-PSDB: AAK63229.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis.

Claim 11: SEQ ID NO 18041; 3071pp + Sequence Listing; English.

AX54951 to AAK64702 encode the human immune/hematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patient's own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the

XX PD 25-MAY-2001.
 XX PF 15-NOV-2000; 2000WO-US31162.
 XX PR 19-NOV-1999; 99US-0166A15.
 XX PR 30-JUN-2000; 2000US-0215136.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Ruben SM, Komatsoulis GA, Baker KP, Young PE;
 XX PI WPI: 2001-343793/36.
 XX DR N-PSDB: AAD08511.
 XX PT Isolated nucleic acid molecule encoding a human secreted protein is
 XX PT used in preventing, treating or ameliorating a medical condition .
 XX PT Claim 11: Page 453; 509pp; English.
 XX PS AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
 CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.
 CC AAE04240-AAE04297 represent human secreted protein fragments or variants.
 CC The secreted proteins and their genes are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 18 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin ageing due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein of the invention.
 XX Sequence 123 AA;

Query Match 55.4%; Score 46; DB 22; Length 123;
 Best Local Similarity 57.1%; Pred. No. 12;
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 AWOLAKOQAQAEK 16
 ||||:||||:
 Db 90 awqlrqgkqgeqr 103

RESULT 10
 AAB94691
 ID AAB94691 standard; Protein; 595 AA.
 XX AAB94691;
 AC AAB94691;

XX 26-JUN-2001 (first entry)
 DT Human protein sequence SEQ ID NO:15659.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.
 XX PN EP1074617-A2.
 XX PD 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-0116126.
 XX PR 29-JUL-1999; 99JP-0248036.
 XX PR 27-AUG-1999; 99JP-0300253.
 XX PR 11-JAN-2000; 2000JP-0118776.
 XX PR 02-MAY-2000; 2000JP-0183767.
 XX PR 09-JUN-2000; 2000JP-0241899.
 XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WI: 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
 CC full-length cDNAs defined in the specification, and for the detection
 CC and/or diagnosis of the abnormality of the proteins encoded by the
 CC full-length cDNAs .

XX PS Claim 8; SEQ ID 15659; 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides, and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 595 AA;

Query Match 53.0%; Score 44; DB 22; Length 595;
 Best Local Similarity 69.2%; Pred. No. 1.2e+02;
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 QLAKOQAQAEKL 17
 ||||:||||:
 Db 581 qlqkqksmeael 593

RESULT 11
 AAY32158
 ID AAY32158 standard; Protein; 641 AA.
 XX AAY32158;
 AC AAY32158;

XX 01-FEB-2000 (first entry)
 DT Human SH3D1A protein.

XX Human SH3D1A protein.

XX SH3D1A; human; Down's syndrome; leukaemia; cancer;
 DE megakaryocytic abnormality; myeloproliferative disorder;
 XX platelet disorder; neural disorder; thrombocytopoenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;
 XX lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.
 XX Homo sapiens.
 JS WO9953062-A2.
 PN 21-OCT-1999.
 PD 16-APR-1999; 99WO-US08371.
 PF 16-APR-1998; 98US-0082007.
 PR (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 XX Korenberg JR, Chen X;
 PI WPI; 1999-633829/54.
 XX N-PSDB; AAZ34574.
 DR Nucleic acid from the human SH3D1A gene and its products, useful for
 XX the diagnosis and treatment of myeloproliferative disorders and
 XX leukaemia.
 XX Claim 14; Fig 15; 99pp; English.
 XX This sequence represents the protein encoded by the human SH3D1A
 XX cDNA clone 9 (see AAZ34574). SH3D1A contributes to the development
 XX of platelets and the pathogenesis of leukaemias, both in general
 XX and in particular those involving the megakaryocytic lineage. The
 XX gene maps to the small candidate region for low platelets on
 XX chromosome 21. The protein includes SH3 domains and EH domains,
 XX both associated with protein-protein interactions and the latter
 XX with maintenance of the cytoskeleton. At least 3 isoforms of
 XX SH3D1A exist (see AAY32157-58). The invention provides methods for
 XX the diagnosis and treatment of megakaryocytic abnormality,
 XX myeloproliferative disorder, platelet disorder, acute leukaemia,
 XX neural disorders, thrombocytopoenia, platelet disorder on
 XX chromosome 21, low platelets in deletion for 21, association of
 XX gains in chromosome 21 with leukaemias, neural abnormalities,
 XX dysfunctions and disorders including brain malformations and
 XX corresponding cognitive dysfunctions, microcephaly, lissencephaly,
 XX and colpocephaly. Methods are also provided: for suppressing
 XX cells unable to regulate themselves; screening for a somatic
 XX alteration in the SH3D1A gene; monitoring the progress and
 XX adequacy of a treatment; monitoring tumour risk progress or
 XX megakaryocytic abnormality; myeloproliferative disorder, or
 XX haematopoietic disorder, platelet disorder or leukaemia; and
 XX screening of drugs for cancer therapy.
 XX Sequence 641 AA;

Query Match 53.0%; Score 44; DB 20; Length 641;
 Best Local Similarity 69.2%; Pred. No. 1.3e+02;
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Y 5 QLAKCKAQAEKEL 17
 |||||: |||:
 b 115 qlqkqksmeaerl 127

RESULT 12
 AAM93229
 ID AAM933229 standard; Protein; 642 AA.
 XX AAM933229;
 XX 06-NOV-2001 (first entry)

XX Human polypeptide, SEQ ID NO: 2647.
 DE Human; full length cDNA; cDNA synthesis; oligo-capping.
 KW Homo sapiens.
 OS EP1130094-A2.
 PN 05-SEP-2001.
 PD 07-JUL-2000; 2000EP-0114089.
 PF 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 DR WPI; 2001-524255/58.
 DR N-PSDB; AAK94139.
 XX 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX Claim 8; SEQ ID NO 2647; 1380pp + sequence listing; English.
 XX The invention relates to primers for synthesising full length cDNA
 XX clones. 830 cDNA molecules encoding a human protein have been
 XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 XX molecules have been determined. Primers for synthesising the full length
 XX cDNA are useful for clarifying the function of the protein encoded by
 XX the cDNA. The full length clones were obtained by construction of full
 XX length enriched cDNA libraries that were synthesised by the oligo-capping
 XX method. The primers enable the production of the full length cDNA easily
 XX without any special methods. The present sequence is a polypeptide
 XX encoded by a full length human cDNA of the invention.
 XX Note: The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in CD-ROM format directly from EPO.
 XX Sequence 642 AA;

Query Match 53.0%; Score 44; DB 22; Length 642;
 Best Local Similarity 69.2%; Pred. No. 1.3e+02;
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 QLAKCKAQAEKEL 17
 |||||: |||:
 Db 618 qlqkqksmeaerl 630

RESULT 13
 AAY32157
 ID AAY32157 standard; Protein; 648 AA.
 XX AAY32157;
 XX 01-FEB-2000 (first entry)
 XX Human SH3D1A protein.
 XX SH3D1A; human; Down's syndrome; leukaemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neural disorder; thrombocytopoenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.
 XX Homo sapiens.

PN WJ9953062-A2.
 XX 21-OCT-1999.
 PD 16-APR-1999; 99WO-US08371.
 XX 16-APR-1998; 98US-0082007.
 PF 16-APR-1998; 98US-0082007.
 PR 16-APR-1998; 98US-0082007.
 XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 PA Korenberg JR, Chen X;
 PI WPI; 1999-633829/54.
 XX N-PSDB: AA234573.
 DR Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia.
 XX Claim 14; Fig 13; 99pp; English.
 PS This sequence represents the protein encoded by the human SH3D1A
 CC cDNA clone 5 (see AA234573). SH3D1A contributes to the development
 CC of platelets and the pathogenesis of leukaemias, both in general
 CC and in particular those involving the megakaryocytic lineage. The
 CC gene maps to the small candidate region for low platelets on
 CC chromosome 21. The protein includes SH3 domains and EH domains,
 CC both associated with protein-protein interactions and the latter
 CC with maintenance of the cytoskeleton. At least 3 isoforms of
 CC SH3D1A exist (see AAY32154-58). The invention provides methods for
 CC the diagnosis and treatment of megakaryocytic abnormality.
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,
 CC neural disorders, thrombocytopenia, platelet disorder, disorder on
 CC chromosome 21, low platelets in deletion for 21, association of
 CC gains in chromosome 21 with leukaemias, neural abnormalities,
 CC dysfunctions and disorders including brain malformations and
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,
 CC and colpocephaly. Methods are also provided: for suppressing
 CC cells unable to regulate themselves; screening for a somatic
 CC alteration of a treatment; monitoring the progress and
 CC adequacy of a treatment; monitoring tumour risk progress or
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC screening of drugs for cancer therapy.
 XX Sequence 648 AA;
 SQ

Query Match 53.0%; Score 44; DB 20; Length 648;
 Best Local Similarity 69.2%; Pred. No. 1.9e+02;
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 5 QLAKKQKQAEKEL 17
 II III: III: I
 Db 618 qlqkqksmeaerl 630

RESULT 14
 ABG20575
 ID ABG20575 standard; Protein: 932 AA.
 XX AC ABG20575;
 XX 13-FEB-2002 (first entry)
 XX Novel human diagnostic protein #20566.
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
 TW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 XX WO200175067-A2.

XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 XX N-PSDB: AAS84762.
 DR New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 20; SEQ ID NO 50934; 103pp; English.
 PS The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 932 AA;
 SQ

Query Match 53.0%; Score 44; DB 22; Length 932;
 Best Local Similarity 69.2%; Pred. No. 1.9e+02;
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 5 QLAKKQKQAEKEL 17
 II III: III: I
 Db 158 qlqkqksmeaerl 170

RESULT 15
 AAM43519
 ID AAM43519 standard; Protein: 1035 AA.
 XX AC AAM43519;
 XX 22-OCT-2001 (first entry)
 XX Human polypeptide SEQ ID NO 197.
 XX Human: antiarthritic; antirheumatic; antiproliferative; vasotropic;
 KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
 KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
 KW antiinflammatory; antiulcer; vulnerary; anticonvulsant; antibacterial;
 KW antiparasitic; cardiant; gene therapy; cancer; immune disorder;
 KW cardiovascular disorder; neurological disease; infection; human.

PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-488781/53.
 DR N-PSDB; AAI63825.
 XX
 PT New isolated nucleic acids and polypeptides, useful for diagnosing,
 PT treating and/or preventing human diseases and disorders -
 XX
 PS Claim 11: SEQ ID NO 197: 654pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI63803-AAI64012) and
 CC the encoded proteins (AAM434497-AAM43660) useful for preventing, treating
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The
 CC genes were isolated from a range of human tissues disclosed in the
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1035 AA;

Query Match 53.0%; Score 44; DB 22; Length 1035;
 Best Local Similarity 69.2%; Pred. No. 2.2e+02;
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

2Y 5 QLAKOKAQAEEKL 17
 |||||
 3b 629 gllqkqksmeaerl 641

Search completed: August 6, 2002, 16:52:53
 Job time: 335 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: August 6, 2002, 16:53:54 ; Search time 51.22 Seconds
(without alignments)
8.107 Million cell updates

Title: US-10-020-139-2_COPY_68_84
 perfect score: 83
 Sequence: 1 SSAWQLAKQKAQAEKL 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
```

Database : Issued Patents AA:*

```

issued_files_in/
1: /cgn2_6/pdatata/2/iaa/5A_COMB.pcp: *
2: /cgn2_6/pdatata/2/iaa/5B_COMB.pcp: *
3: /cgn2_6/pdatata/2/iaa/6A_COMB.pcp: *
4: /cgn2_6/pdatata/2/iaa/6B_COMB.pcp: *
5: /cgn2_6/pdatata/2/iaa/PCRTUS_COMB.pcp: *
6: /cgn2_6/pdatata/2/iaa/backfiles1.pcp: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query		DB	ID	Description
		Match	%			
1	42	50.6	555	2	US-08-793-229-34	Sequence 34, Appl
2	42	50.6	555	3	US-08-793-957-34	Sequence 34, Appl
3	42	50.6	557	2	US-08-793-229-33	Sequence 33, Appl
4	42	50.6	557	3	US-09-285-957-33	Sequence 33, Appl
5	42	50.6	844	1	US-07-731-157A-6	Sequence 6, Appl
6	42	50.6	844	2	US-08-541-780-6	Sequence 6, Appl
7	42	50.6	846	1	US-07-731-157A-5	Sequence 5, Appl
8	42	50.6	846	2	US-08-541-780-5	Sequence 5, Appl
9	39.5	47.6	40	2	US-08-557-309B-27	Sequence 27, Appl
10	39.5	47.6	40	3	US-08-834-306-27	Sequence 27, Appl
11	39.5	47.6	40	4	US-08-993-674A-27	Sequence 27, Appl
12	39.5	47.6	186	2	US-08-557-309B-43	Sequence 43, Appl
13	39.5	47.6	186	3	US-08-834-306-43	Sequence 43, Appl
14	39.5	47.6	186	4	US-08-993-674A-43	Sequence 43, Appl
15	38.5	46.4	734	4	US-09-185-160-13	Sequence 13, Appl
16	38	45.8	280	2	US-08-424-641B-12	Sequence 12, Appl
17	38	45.8	280	2	US-08-820-980-12	Sequence 12, Appl
18	38	45.8	280	2	US-08-826-439-12	Sequence 12, Appl
19	38	45.8	1507	6	5268270-2	Patent No. 5268270
20	37.5	45.2	731	4	US-09-185-160-11	Sequence 11, Appl
21	37	44.6	155	1	US-08-530-010-13	Sequence 13, Appl
22	37	44.6	155	2	US-08-484-101B-13	Sequence 13, Appl
23	37	44.6	155	4	US-08-714-524D-13	Sequence 13, Appl
24	37	44.6	246	4	US-09-185-160-7	Sequence 7, Appl
25	37	44.6	641	4	US-08-961-083-160	Sequence 160, App
26	37	44.6	918	2	US-08-843-530B-35	Sequence 35, Appl
27	36	43.4	91	2	US-08-997-080-143	Sequence 143, App

Query Match 50.6%; Score 42; DB 2; Length 555;
Best Local Similarity 41.2%; Pred. No. 64;
Matches 7; Conservative 5; Mismatches 5; Indels

QY 1 SSAWQLAKQKAQAEKL 17
| : | : | | | : | : | : |

Sequence 143, App
Sequence 143, App
Sequence 143, App
Sequence 102, App
Sequence 196, App
Sequence 21, Appl
Sequence 9, Appl
Sequence 6, Appl
Sequence 76, Appl
Sequence 7, Appl
Sequence 7, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 395, App
Sequence 2, Appl
Sequence 4, Appl

ALIGNMENTS

```

1
RESULT
US-08-793-229-34
; Sequence 34, Application US/08793229
; Patent No. 5891703
; GENERAL INFORMATION:
; APPLICANT: VAN DER LAAN, Jan Metske
; APPLICANT: RIEMENS, Adriana Marina
; APPLICANT: QUAX, Wilhelmus Johannes
; TITLE OF INVENTION: Mutated Penicillin G Acylase Genes
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,229
; FILING DATE: 23-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/03249
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 97075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)913-0001
; TELEFAX: (312)913-0002
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-793-229-34

```

0; Gaps 0;

b 1 SNMWIGKNAQADAKAI 17

RESULT 2

S-09-285-957-34
Sequence 34, Application US/09285957
Patent No. 6033823
GENERAL INFORMATION:
APPLICANT: VAN DER LAAN, Jan Metske
APPLICANT: RIEMENS, Adriana Marina
APPLICANT: QUAX, Wilhelmus Johannes
TITLE OF INVENTION: Mutated Penicillin G Acylase Genes
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/285.957
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/793,229
FILING DATE:
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)913-0001
TELEFAX: (312)913-0002
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 555 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
S-09-285-957-34

Query Match 50.6%; Score 42; DB 3; Length 555;
Best Local Similarity 41.2%; Pred. No. 64;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Y 1 SSAWQLAKQKQAEAKL 17
I: | | | | | | | | | |
Z 1 SNMWIGKNAQADAKAI 17

RESULT 3

S-08-793-229-33
Sequence 33, Application US/08793229
Patent No. 5891703
GENERAL INFORMATION:
APPLICANT: VAN DER LAAN, Jan Metske
APPLICANT: RIEMENS, Adriana Marina
APPLICANT: QUAX, Wilhelmus Johannes
TITLE OF INVENTION: Mutated Penicillin G Acylase Genes
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,229
FILING DATE: 23-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/03249
FILING DATE:
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)913-0001
TELEFAX: (312)913-0002
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 557 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-793-229-33

Query Match 50.6%; Score 42; DB 2; Length 557;
Best Local Similarity 41.2%; Pred. No. 64;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 SSAWQLAKQKQAEAKL 17
I: | | | | | | | | | |
DB 1 SNMWIGKNAQADAKAI 17

RESULT 4

US-09-285-957-33
Sequence 33, Application US/09285957
Patent No. 6033823
GENERAL INFORMATION:
APPLICANT: VAN DER LAAN, Jan Metske
APPLICANT: RIEMENS, Adriana Marina
APPLICANT: QUAX, Wilhelmus Johannes
TITLE OF INVENTION: Mutated Penicillin G Acylase Genes
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/285.957
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/793,229
FILING DATE:
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)913-0001
TELEFAX: (312)913-0002
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 557 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-285-957-33

Query Match 50.6%; Score 42; DB 3; Length 844;
Best Local Similarity 41.2%; Pred. No. 98;
Matches 7; Conservative 5; Mismatches 5; Indels 5; Gaps 0;

QY 1 SSAMOLAKQKQAEKL 17
Db 1 SNMWIGSKAQDAKAI 17

RESULT 5

US-07-731-157A-6
Sequence 6, Application US/07731157A
Patent No. 5457032

GENERAL INFORMATION:

APPLICANT: Quax, Wilhelmus J.
APPLICANT: Misset, Onno
APPLICANT: Van der Laan, Jan M.
APPLICANT: Lenting, Herman B.M.
TITLE OF INVENTION: Mutated beta-lactam acylase genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07731.157A
FILING DATE: 19910509

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 90200962

FILING DATE: 18-APR-1990

ATTORNEY/AGENT INFORMATION:

NAME: RAE-VENTER PH.D., BARBARA

REGISTRATION NUMBER: 32,750

REFERENCE/DOCKET NUMBER: GBRO-027/0005

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-494-7622

TELEFAX: 415-857-0663

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 844 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORGANISM: Kluyvera citrophila

STRAIN: ATCC 21285

US-07-731-157A-6

Query Match 50.6%; Score 42; DB 1; Length 844;
Best Local Similarity 41.2%; Pred. No. 98;
Matches 7; Conservative 5; Mismatches 5; Indels 5; Gaps 0;

QY 1 SSAMOLAKQKQAEKL 17
Db 1 SNMWIGSKAQDAKAI 17

Db 290 SNMWIGKNKQADAKAI 306

RESULT 6

US-08-541-780-6
Sequence 6, Application US/08541780
Patent No. 5935831

GENERAL INFORMATION:

APPLICANT: Quax, Wilhelmus J.
APPLICANT: Misset, Onno
APPLICANT: Van der Laan, Jan M.
APPLICANT: Lenting, Herman B.M.
TITLE OF INVENTION: Mutated beta-lactam acylase genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/541.780
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/731.157

FILING DATE:

APPLICATION NUMBER: EP 90200962

FILING DATE: 18-APR-1990

ATTORNEY/AGENT INFORMATION:

NAME: RAE-VENTER PH.D., BARBARA

REGISTRATION NUMBER: 32,750

REFERENCE/DOCKET NUMBER: GBRO-027/0005

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-494-7622

TELEFAX: 415-857-0663

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 844 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORGANISM: Kluyvera citrophila

STRAIN: ATCC 21285

US-08-541-780-6

Query Match 50.6%; Score 42; DB 2; Length 844;
Best Local Similarity 41.2%; Pred. No. 98;
Matches 7; Conservative 5; Mismatches 5; Indels 5; Gaps 0;

QY 1 SSAMOLAKQKQAEKL 17

Db 290 SNMWIGKNKQADAKAI 306

RESULT 7

US-07-731-157A-5
Sequence 5, Application US/07731157A
Patent No. 5457032

GENERAL INFORMATION:

APPLICANT: Quax, Wilhelmus J.
APPLICANT: Misset, Onno
APPLICANT: Van der Laan, Jan M.

APPLICANT: Lenting, Herman B.M.
 TITLE OF INVENTION: Mutated beta-lactam acylase genes
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
 STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
 CITY: PALO ALTO
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 94306
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/731.157A
 FILING DATE: 19910509
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: EP 90200962
 FILING DATE: 18-APR-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: RAE-VENTER PH.D., BARBARA
 REGISTRATION NUMBER: 32,750
 REFERENCE/DOCKET NUMBER: GBRO-027/0005
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-494-7622
 TELEFAX: 415-857-0663
 TELEX: 380816 COOLEY PA
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 846 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Escherichia coli
 PUBLICATION INFORMATION:
 AUTHORS: Schumacher, G
 AUTHORS: Sizmann, D
 AUTHORS: Haug, H
 AUTHORS: Buckel, P
 AUTHORS: Bock, A
 TITLE: Penicillin acylase from E.coli: unique
 TITLE: gene-protein reaction.
 JOURNAL: Nucleic Acids Res.
 VOLUME: 14
 PAGES: 5713-5727
 DATE: 1986
 S-07-731-157A-5

Query Match 50.6%; Score 42; DB 1; Length 846;
 Best Local Similarity 41.2%; Pred. No. 98;
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Y 1 SSAWOLAKOAEAEKL 17
 | : | : | : | : | :
 D 290 SNMWVIGSKAQDAKAI 306

RESULT 8
 S-08-541-780-5
 Sequence 5, Application US/08541780
 Patent No. 5935831
 GENERAL INFORMATION:
 APPLICANT: Quax, Wilhelmus J.
 APPLICANT: Misset, Onno
 APPLICANT: Van der Laan, Jan M.
 APPLICANT: Lenting, Herman B.M.
 TITLE OF INVENTION: Mutated beta-lactam acylase genes
 NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:
 ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
 STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
 CITY: PALO ALTO
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 94306
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/541,780
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/731,157
 FILING DATE:
 APPLICATION NUMBER: EP 90200962
 FILING DATE: 18-APR-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: RAE-VENTER PH.D., BARBARA
 REGISTRATION NUMBER: 32,750
 REFERENCE/DOCKET NUMBER: GBRO-027/0005
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-494-7622
 TELEFAX: 415-857-0663
 TELEX: 380816 COOLEY PA
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 846 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Escherichia coli
 PUBLICATION INFORMATION:
 AUTHORS: Schumacher, G
 AUTHORS: Sizmann, D
 AUTHORS: Haug, H
 AUTHORS: Buckel, P
 AUTHORS: Bock, A
 TITLE: Penicillin acylase from E.coli: unique
 TITLE: gene-protein reaction.
 JOURNAL: Nucleic Acids Res.
 VOLUME: 14
 PAGES: 5713-5727
 DATE: 1986
 US-08-541-780-5

Query Match 50.6%; Score 42; DB 2; Length 846;
 Best Local Similarity 41.2%; Pred. No. 98;
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 SSAWOLAKOAEAEKL 17
 | : | : | : | : | :
 DB 290 SNMWVIGSKAQDAKAI 306

RESULT 9
 US-08-557-309B-27
 Sequence 27, Application US/08557309B
 Patent No. 5916572
 GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Lodes, Michael J.
 APPLICANT: Houghton, Raymond L.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF
 NUMBER OF SEQUENCES: 69
 CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557.309B
FILING DATE: 14-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-557-309B-27

Query Match 47.6%; Score 39.5; DB 2; Length 40;
Best Local Similarity 60.0%; Pred. No. 10;
Matches 9; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 3 AWOLAKORAO-EAEK 16
DB 22 AWQAEAEAEAEAE 36

RESULT 10
US-08-834-306-27
Sequence 27, Application US/08834306
Patent No. 6054135
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834.306
FILING DATE: 15-APR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-834-306-27

Query Match 47.6%; Score 39.5; DB 3; Length 40;
Best Local Similarity 60.0%; Pred. No. 10;
Matches 9; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 3 AWOLAKORAO-EAEK 16
DB 22 AWQAEAEAEAEAE 36

RESULT 11
US-08-993-674A-27
Sequence 27, Application US/08993674A
Patent No. 6228372
GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993.674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-993-674A-27

Query Match 47.6%; Score 39.5; DB 4; Length 40;
Best Local Similarity 60.0%; Pred. No. 10;
Matches 9; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 3 AWOLAKORAO-EAEK 16
DB 22 AWQAEAEAEAEAE 36

```

RESULT 12
S-08-557-309B-43
Sequence 43, Application US/08557309B
Patent No. 5916572
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
SOFTWARE: PatentIn Release #1.0, Version #1.30
APPLICATION NUMBER: US/08/557.309B
FILING DATE: 14-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
S-08-557-309B-43

Query Match 47.6%; Score 39.5; DB 2; Length 186;
Best Local Similarity 60.0%; Pred No. 50;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

Y 3 AWOLAKOKAQ-EAEK 16
||| |::||| |||
c 111 AWQAEERAEAE 125

RESULT 13
S-08-834-306-43
Sequence 43, Application US/08834306
Patent No. 6054135
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
SOFTWARE: PatentIn Release #1.0, Version #1.30
APPLICATION NUMBER: US/08/993.674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
S-08-834-306-43

Query Match 47.6%; Score 39.5; DB 3; Length 186;
Best Local Similarity 60.0%; Pred No. 50;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 3 AWOLAKOKAQ-EAEK 16
||| |::||| |||
Db 111 AWQAEERAEAE 125

RESULT 14
US-08-993-674A-43
Sequence 43, Application US/08993674A
Patent No. 6228372
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John W.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
SOFTWARE: PatentIn Release #1.0, Version #1.30
APPLICATION NUMBER: US/08/993.674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-834-306-43

```

US-08-993-674A-43

Query Match 47.6%; Score 39.5; DB 4; Length 186;
Best Local Similarity 60.0%; Pred. No. 50;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 3 AWOLAKAKAQ-AEK 16
||| |::||| |||
Db 111 AWQAEERAKQAEAE 125

RESULT 15
US-09-185-160-13
; Sequence 13, Application US/09185160
; Patent No. 6252137
; GENERAL INFORMATION:
; APPLICANT: ODELL, JOAN T.
; TITLE OF INVENTION: SOYBEAN HOMOLOG OF A SEED-SPECIFIC
; TITLE OF INVENTION: TRANSCRIPTION ACTIVATOR FROM PHASEOLUS
; TITLE OF INVENTION: VULGARIS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/185,160
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/065,459
; FILING DATE: NOVEMBER 12, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: MAJARIAN, WILLIAM R.
; REGISTRATION NUMBER: 41,173
; REFERENCE/DOCKET NUMBER: BB-1096
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4926
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 734 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
JS-09-185-160-13

Query Match 46.4%; Score 38.5; DB 4; Length 734;
Best Local Similarity 52.9%; Pred. No. 2.9e+02;
Matches 9; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

ZY 1 SSANOLAKAKAQ-AEK 16
||| |::||| |||
Db 76 SSSWMLKSDAEAEK 92

Search completed: August 6, 2002, 16:53:55
Job time: 287 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

Ⓜ protein - protein search, using sw model

run on: August 6, 2002, 17:05:15 ; Search time 66.51 Seconds

(without alignments)
13.003 Million cell updates/sec

title: US-10-020-139-2_COPY_115_123

sequence score: 49

1 EPIDCKGL 9

scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 283138 seqs, 96089334 residues

total number of hits satisfying chosen parameters: 283138

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

database :

PIR71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	44	89.8	359	2 AI0641	fatty acid/phospholipid
2	40	81.6	413	2 T21644	hypothetical prote
3	38	77.6	627	2 T10290	hypothetical prote
4	37	75.5	428	2 AE2118	carboxyl-terminal
5	37	75.5	766	2 C87689	NADP-dependent mal
6	36	73.5	456	2 A83218	conserved hypothet
7	36	73.5	470	2 S30597	H+-transporting AT
8	36	73.5	744	2 T35192	probable ABC trans
9	36	73.5	5627	2 C83339	hypothetical prote
10	35	71.4	152	2 D69989	hypothetical prote
11	35	71.4	387	2 D70090	hypothetical prote
12	35	71.4	397	2 AF3134	hydrolyase [impor
13	35	71.4	400	2 F98153	N-isopropylammelid
14	35	71.4	614	2 T16268	hypothetical prote
15	35	71.4	626	2 G82078	topoisomerase IV
16	35	71.4	629	2 B83026	topoisomerase IV s
17	35	71.4	629	2 T43812	topoisomerase IV c
18	35	71.4	630	2 D65090	topoisomerase IV s
19	35	71.4	630	2 S33711	DNA topoisomerase
20	35	71.4	630	2 F91118	DNA topoisomerase
21	35	71.4	630	2 E85963	DNA topoisomerase
22	35	71.4	630	2 A80889	topoisomerase IV c
23	35	71.4	631	2 AG0082	topoisomerase IV c
24	35	71.4	2910	2 T28156	DNA-directed RNA p
25	34	69.4	132	2 B64474	hypothetical prote
26	34	69.4	282	2 B99960	D-alanine aminotra
27	34	69.4	323	2 C56024	GDP dissociation i
28	34	69.4	367	2 C83815	hypothetical prote
29	34	69.4	428	2 S15662	farnesyltransferas

ALIGNMENTS

RESULT 1

AI0641

fatty acid/phospholipid synthesis protein PlsX [imported] - Salmonella enterica subs
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AI0641

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Far, S.; Moulé, S.; O'Gaora, P.
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, A.; Title: Complete genome sequence of a multiple drug resistant Salmonella enterica s

A:Reference number: AB0502; PMID:11677608

A:Accession: AI0641

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-359 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08316.1; PID:gi16502362; GSPDB:GN00176

C:Genetics:

A:Gene: STY1231

C:Superfamily: phospholipid synthesis protein

Query Match 89.8%; Score 44; DB 2; Length 359;
Best Local Similarity 77.8%; Pred. No. 0.99;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIDCKGL 9

DB 345 EPIDCKGLV 353

|||||

RESULT 2

T21644

hypothetical protein F32B6.9 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T21644

R:Basham, V.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19453

A:Accession: T21644

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-413 <WIL>

A:Cross-references: EMBL:Z81074; PIDN:CAB03043.1; GSPDB:GN00022; CESP:F32B6.9

C:Experimental source: clone F32B6

C:Genetics:

A:Gene: CESP:F32B6.9

A:Map position: 4

A:Introns: 22/2; 61/1; 141/2; 176/3; 200/3; 290/3; 332/1

GDP-dissociation i
smg p25A regulator
GDP dissociation i
rab GDP dissociati
GDP dissociation i
YUP812R.39, homol
hypothetical prote
earl protein - mal
protein K10D2.6 [i
mitogen-activated
chemotaxis protein
tail tip fiber pro
hypothetical prote
hypothetical prote
hypothetical prote

30 34 69.4 447 2 I37082
31 34 69.4 447 2 A35852
32 34 69.4 447 2 B56024
33 34 69.4 447 2 A54091
34 34 69.4 448 2 S36746
35 34 69.4 464 2 G86453
36 34 69.4 586 2 T08293
37 34 69.4 656 2 T01573
38 34 69.4 662 2 C88451
39 34 69.4 721 2 A56352
40 34 69.4 723 2 AD0203
41 34 69.4 769 2 B81447
42 34 69.4 1061 2 T13107
43 34 69.4 1139 2 B70954
44 34 69.4 3263 2 E82410
45 33 67.3 134 2 D97401

Query Match 81.6%; Score 40; DB 2; Length 413;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 PIDDGKG 8
 |||||
 384 PIDDGKG 390

RESULT 3
 10290
 /pothetical protein 21 - Orgyia pseudotsugata nuclear polyhedrosis virus
 :Species: Orgyia pseudotsugata nuclear polyhedrosis virus, OpMNPV
 :Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 :Accession: T10290
 :Ahrens, C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohrmann, G.F.
 :Virology 229, 381-399, 1997
 :Title: The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedrosis
 :Reference number: 217011; MUID:97271300
 :Accession: AE2118
 :Status: preliminary; translated from GB/EMBL/DBJ
 :Molecule type: DNA
 :Residues: 1-627 <AHR>
 :Cross-references: EMBL:U75930; NID:g2934903; PIDN:AAC59020.1; PID:g1911267

Query Match 77.6%; Score 38; DB 2; Length 627;
 Best Local Similarity 87.5%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2 PIDDGKGL 9
 |||||
 25 PIDDGAGL 32

RESULT 4
 2118
 rboxyl-terminal proteinase [imported] - Anabaena sp. (strain PCC 7120)
 :Species: Anabaena sp.
 :Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
 :Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
 :Accession: AE2118
 :Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 :akazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 :A Res. 8, 205-213, 2001
 :Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 :Reference number: AB1807; MUID:21595285; PMID:11759840
 :Accession: AE2118
 :Status: preliminary
 :Molecule type: DNA
 :Residues: 1-428 <KUK>
 :Cross-references: GB:BA000019; PIDN:BA074199.1; PID:g17131592; GSPDB:GN00179
 :Experimental source: strain PCC 7120
 :Genetics:
 :Gene: al12500
 :Superfamily: carboxyl-terminal processing proteinase

Query Match 75.5%; Score 37; DB 2; Length 428;
 Best Local Similarity 75.0%; Pred. No. 28;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

2 PIDDGKGL 9
 |||||
 347 PLDDGSGL 354

RESULT 5
 7689
 2P-dependent malic enzyme [imported] - Caulobacter crescentus
 :Species: Caulobacter crescentus

C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: C87689
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
 n. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.
 :Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: C87689
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-766 <STO>
 A:Cross-references: GB:AE005673; NID:q13425285; PIDN:AAK25511.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC3549

Query Match 75.5%; Score 37; DB 2; Length 766;
 Best Local Similarity 62.5%; Pred. No. 51;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIDGKG 8
 :||:|:|:|
 DB 759 QPVDGEG 766

RESULT 6
 A83218
 conserved hypothetical protein PA3421 [imported] - Pseudomonas aeruginosa (strain PA
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: AB3218
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.
 :adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
 :Lory, S.; Olson, M.V.
 :Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
 A:Reference number: A82950; MUID:20437337
 A:Accession: AB3218
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-456 <STO>
 A:Cross-references: GB:AE004763; GB:AE004091; NID:g9949556; PIDN:AAG06809.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA3421

Query Match 73.8%; Score 36; DB 2; Length 456;
 Best Local Similarity 66.7%; Pred. No. 46;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EPIDGKGL 9
 :||:|:|
 DB 141 QPINDGNGL 149

RESULT 7
 S30597
 H+-transporting ATP synthase (EC 3.6.1.34) beta chain - Lactobacillus casei
 C:Species: Lactobacillus casei
 C:Date: 22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change 19-Jan-2001
 C:Accession: S30597
 R:Klugbauer, N.; Ludwig, W.; Baeuerlein, E.; Schleifer, K.H.
 :Syst. Appl. Microbiol. 15, 323-330, 1992
 A:Title: Subunit beta of adenosine triphosphate synthase of Pectinatus frisingensis a
 A:Reference number: S30597
 A:Accession: S30597
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-470 <KLJ>
 A:Cross-references: EMBL:X64542; NID:g43971; PIDN:CAA45840.1; PID:g43972
 C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase

;Keywords: ATP biosynthesis; hydrolase; nucleotide binding; P-loop
 ;155-162/Region: nucleotide-binding motif A (P-loop)
 ;181-350/Domain: H-tt-transporting ATP synthase alpha chain homology <ATP>

Query Match 73.5%; Score 36; DB 2; Length 470;
 Best Local Similarity 66.7%; Pred. No. 48;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Y 1 EPIDDDGKGL 9
 :|||||:
 b 98 DPIDDDGKGL 106

RESULT 8

robable ABC transporter - Streptomyces coelicolor
 ;Species: Streptomyces coelicolor
 ;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 ;Accession: T35192

;Seeger, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL data Library, April 1998

;Reference number: Z21571

;Accession: T35192

;Status: preliminary; translated from GB/EMBL/DDBJ

;Molecule type: DNA

;Residues: 1-744 <SEE>

;Cross-references: EMBL:AL022374; PIDN:CAA18516.1; GSPDB:GN000070; SCOEDB:SC5B8.08

;Experimental source: strain A3(2)

;Genetics:

;Gene: SCOEDB:SC5B8.08

Query Match 73.5%; Score 36; DB 2; Length 744;
 Best Local Similarity 62.5%; Pred. No. 78;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Y 1 EPIDDDGK 8
 :|||||:
 b 341 EPLEDDGK 348

RESULT 9

hypothetical protein PA2462 [imported] - Pseudomonas aeruginosa (strain PA01)

;Species: Pseudomonas aeruginosa

;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

;Accession: C83339

;Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B

dan, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

ature 406, 959-964, 2000

;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path

;Reference number: A82950; MUID:20437337

;Accession: C83339

;Status: preliminary

;Molecule type: DNA

;Residues: 1-5627 <STO>

;Cross-references: GB:AE004673; GB:AE004091; NID:g9948501; PIDN:AG05850.1; GSPDB:GN001

;Experimental source: strain PA01

;Genetics:

;Gene: PA2462

Query Match 73.5%; Score 36; DB 2; Length 5627;
 Best Local Similarity 87.5%; Pred. No. 6.7e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 2 PIDDDGKGL 9
 :|||||:
 b 5280 PIDDDGKGL 5287

RESULT 10
 D69989

hypothetical protein yycG - Bacillus subtilis

N;Alternate names: hypothetical protein x (dnaB 5' region)

C;Species: Bacillus subtilis

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C;Accession: D69989; A26580

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Ga

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin,

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma

Y, M.; Ogawa, K.; Ogilwa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scan

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiy

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshid

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil

A;Reference number: A69580; MUID:98044033

A;Accession: D69989

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-152 <KUN>

A;Cross-references: GB:Z99118; GB:AL009126; NID:g2635200; PIDN:CAB14860.1; PID:g2635

A;Experimental source: strain 168

R;Hoshino, T.; McKenzie, T.; Schmidt, S.; Tanaka, T.; Sueoka, N.

Proc. Natl. Acad. Sci. U.S.A. 84, 653-657, 1987

A;Title: Nucleotide sequence of Bacillus subtilis dnaB: a gene essential for DNA rep

A;Reference number: A94709; MUID:87118226

A;Accession: A26580

A;Molecule type: DNA

A;Residues: 60-152 <HOS>

A;Cross-references: GB:M15183; NID:g142862; PIDN:AAA22403.1; PID:g468267

C;Genetics:

A;Gene: yycG

C;Superfamily: conserved hypothetical protein HI0943

Query Match 71.4%; Score 35; DB 2; Length 152;
 Best Local Similarity 62.5%; Pred. No. 23;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 PIDDGKGL 9

:|||||:

Db 18 PVDDGKSI 25

RESULT 11

D70090

hypothetical protein yycP - Bacillus subtilis

C;Species: Bacillus subtilis

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C;Accession: D70090

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Ga

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin,

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma

Y, M.; Ogawa, K.; Ogilwa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scan

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiy

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshid

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil

A;Reference number: A69580; MUID:98044033

Accession: D70090
 Status: preliminary; nucleic acid sequence not shown; translation not shown
 Molecule type: DNA
 Residues: 1-387 <KUN>
 Cross-references: GB:299124; GB:AL009126; NID:g2636442; PIDN:CA816064.1; PID:g2636574
 Experimental source: strain 168
 Genetics:
 Gene: yycP
 Superfamily: Bacillus subtilis hypothetical protein yycP

Query Match 71.4%; Score 35; DB 2; Length 387;
 Best Local Similarity 85.7%; Pred. No. 61;
 Matches 6; Conservative 0; Mismatches 0; Gaps 0;
 2 PIDGKG 8
 185 PTDGKG 191

RESULT 12
 3134
 drolase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 Species: Agrobacterium tumefaciens
 Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 Accession: AF3134
 Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 age, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 er, E.W.

Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

Reference number: AB2577; PMID:11743193

Accession: AF3134

Status: preliminary

Molecule type: DNA

Residues: 1-397 <KUR>

Cross-references: GB:AE008689; PIDN:AAL45492.1; PID:g17743200; GSPDB:GN00187

Experimental source: strain C58 (Dupont)

Genetics:

Gene: Atu4698

Map position: linear chromosome

Query Match 71.4%; Score 35; DB 2; Length 397;
 Best Local Similarity 71.4%; Pred. No. 63;
 Matches 5; Conservative 2; Mismatches 0; Gaps 0;
 2 PIDGKG 8
 39 PVEDGKG 45

RESULT 13
 8153
 isopropylamide isopropyl amidohydrolase [imported] - Agrobacterium tumefaciens (str
 Species: Agrobacterium tumefaciens
 Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
 Accession: F98153
 Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldman,
 ; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 ence 294, 2323-2328, 2001
 Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 Reference number: A97359; PMID:11743194

Accession: F98153

Status: preliminary

Molecule type: DNA

Residues: 1-400 <KUR>

Cross-references: GB:AE007870; PIDN:AAK88752.1; PID:g15158494; GSPDB:GN00170

Genetics:

Gene: AGR_L_364

Map position: linear chromosome

Query Match 71.4%; Score 35; DB 2; Length 400;
 Best Local Similarity 71.4%; Pred. No. 63;
 Matches 5; Conservative 2; Mismatches 0; Gaps 0;
 QY 2 PIDGKG 8
 DB 42 PVEDGKG 48

RESULT 14
 116288
 hypothetical protein F35D11.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16268
 R:Fulton, B.

submitted to the EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid F35D11.

A:Reference number: 218487

A:Accession: T16268

A:Status: preliminary; translated from GB/EMBL/DD8J

A:Molecule type: DNA

A:Residues: 1-614 <FUL>

A:Cross-references: EMBL:U29381; NID:g868214; PID:g868216; PIDN:AAA68749.1; CESP:F35D

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:F35D11.2

A:Introns: 16/2; 105/3; 172/1; 236/2; 341/3; 373/2; 391/2

Query Match 71.4%; Score 35; DB 2; Length 614;
 Best Local Similarity 77.8%; Pred. No. 99;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EPIDGKG 9
 DB 579 EPINDGKFL 587

RESULT 15
 682078
 topoisomerase IV, chain B VC2431 [imported] - Vibrio cholerae (strain N16961 serogrou
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: G82078
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
 I. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: G82078

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-626 <HEI>

A:Cross-references: GB:AE004312; GB:AE003852; NID:g9656995; PIDN:AAF95574.1; GSPDB:GN

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC2431

A:Map position: 1

C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 71.4%; Score 35; DB 2; Length 626;
 Best Local Similarity 66.7%; Pred. No. 1e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EPIDGKG 9
 DB 66 EVIDDGRGM 74

Search completed: August 6, 2002, 17:05:16
Job time: 918 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

protein - protein search, using sw model

on: August 6, 2002, 17:07:35 ; Search time 32.88 Seconds

(without alignments)

10.598 Million cell updates/sec

title: US-10-020-139-2_COPY_115_123

effect score: 49

sequence: 1 EPIDCKGL 9

scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 105224 seqs, 38719550 residues

total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	44	89.8	359	1	PLSX_SALTY
2	40	81.6	413	1	VV4Q_CABEL
3	38	77.6	657	1	Y023_NPVOP
4	36	73.5	470	1	ATPB_LACCA
5	35	71.4	152	1	YTCG_BACSU
6	35	71.4	630	1	PARE_ECOLI
7	35	71.4	630	1	PARE_SALTY
8	34	69.4	323	1	GD1A_MOUSE
9	34	69.4	428	1	GGPP_NEUCR
10	34	69.4	447	1	GD1A_MOUSE
11	34	69.4	447	1	GD1A_MOUSE
12	34	69.4	447	1	GD1A_MOUSE
13	34	69.4	447	1	GD1A_MOUSE
14	34	69.4	721	1	MD06_HUMAN
15	33	67.3	136	1	RL28_HUMAN
16	33	67.3	136	1	RL28_MOUSE
17	33	67.3	192	1	RL28_MOUSE
18	33	67.3	233	1	RNC_COXBU
19	33	67.3	272	1	RL5_BP74
20	33	67.3	299	1	RL5_BOMMO
21	33	67.3	331	1	GSYP_AERYH
22	33	67.3	396	1	XYNA_BACCH
23	33	67.3	404	1	RL17_MOUSE
24	33	67.3	448	1	RL17_MOUSE
25	33	67.3	584	1	EXON_HCMVA
26	33	67.3	603	1	SNTD_HAEIN
27	33	67.3	868	1	ENV_HV1C4
28	33	67.3	1537	1	DNM1_CHICK
29	33	67.3	2151	1	RRPL_HANTY
30	33	67.3	2156	1	RRPL_PUMNH
31	32	65.3	3256	1	KT67_HUMAN
32	32	65.3	266	1	LEFT_NPVAC
33	32	65.3	312	1	HO2_RABIT

34	32	65.3	315	1	HO2_MOUSE
35	32	65.3	315	1	HO2_RAT
36	32	65.3	316	1	HO2_HUMAN
37	32	65.3	356	1	NTRB_RHOCA
38	32	65.3	519	1	DHA4_YEAST
39	32	65.3	609	1	KPK1_PHAVU
40	32	65.3	663	1	SEPI_SCHPO
41	32	65.3	664	1	PD13_HUMAN
42	32	65.3	811	1	YJ96_YEAST
43	32	65.3	819	1	ADVL_MOUSE
44	32	65.3	859	1	YQD3_CAEEL
45	32	65.3	947	1	BGL3_RUMAL

ALIGNMENTS

RESULT 1	PLSX_SALTY	STANDARD:	PRT:	359 AA.
ID	PLSX_SALTY	STANDARD:	PRT:	359 AA.
AC	O85138;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Fatty acid/phospholipid synthesis protein plsx.			
GN	PLSX OR STM1192 OR STY1231.			
OS	Salmonella typhimurium, and			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Salmonella.			
OX	NCBI_Taxid--602, 601;			
RP	SEQUENCE FROM N.A.			
RC	SPECIES-S.typhimurium; STRAIN-LT2;			
RX	MEDLINE=98317265; PubMed=9642179;			
RT	Zhang Y., Cronan J.E. Jr.,			
RT	"Transcriptional analysis of essential genes of the Escherichia coli			
RT	fatty acid biosynthesis gene cluster by functional replacement with			
RT	the analogous Salmonella typhimurium gene cluster.";			
RL	J. Bacteriol. 180:3295-3303(1998).			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES-S.typhimurium; STRAIN-LT2 / SGSC1412 / ATCC 700720;			
RX	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Bante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.";			
RL	Nature 413:852-856(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES-S.typhi; STRAIN-CT18;			
RX	MEDLINE=21534947; PubMed=11677608;			
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,			
RA	Churcher C., Dougan K.L., Bentley S.D., Holden M.T.G., Sebahia M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jajels K.,			
RA	Krogn A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,			
RA	Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Barrell B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhi CT18.";			
RL	Nature 413:848-852(2001).			
CC	!- FUNCTION: NOT KNOWN, PROBABLY INVOLVED IN FATTY ACID OR			
CC	PHOSPHOLIPID SYNTHESIS (BY SIMILARITY).			
CC	!- SIMILARITY: BELONGS TO THE PLSX FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF044668; AAC38647.1; ALT_INIT.
EMBL; AE008731; AAL20121.1; -
EMBL; AL627289; CAD08316.1; -
StyGene; SG10682; p1sx.
InterPro; IPR003664; FA_synthesis.
Pfam; PF02504; FA_synthesis; 1.
Fatty acid biosynthesis; Phospholipid biosynthesis; Complete proteome.
SEQUENCE 359 AA; 38716 MW; ADD01BA971D5ECD8 CRC64;

Query Match 89.8%; Score 44; DB 1; Length 359;
Best Local Similarity 77.8%; Pred. No. 0.31;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

1 EPIDGKGL 9
|||||
345 EPLDGGKV 353

SUIT 2
4Q_CAEEL STANDARD; PRT; 413 AA.
O45435;
15-JUL-1999 (Rel. 38, Created)
15-JUL-1999 (Rel. 38, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Hypothetical 49.0 kDa protein F32B6.9 in chromosome IV.
F32B6.9.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
Rhabditidae; Peloderinae; Caenorhabditis.
NCBI_TaxID=6239;
[1]

SEQUENCE FROM N.A.
STRAIN-BRISTOL N2;
Basham V.;
Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
-!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Z81074; CAB03043.1; -
WormPep; F32B6.9; CE09864.
InterPro; IPR000615; Worm_fam_8.
Pfam; PF01062; Worm_fam_8; 1.
ProDom; PD002802; Worm_fam_8; 1.
Hypothetical protein.
SEQUENCE 413 AA; 48965 MW; A66E69A83C78790B CRC64;

Query Match 81.6%; Score 40; DB 1; Length 413;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 PIDDGKGL 8
|||||
384 PIDDGKGL 390

SUIT 3
23_NPVOP

ID Y023_NPVOP STANDARD; PRT; 657 AA.
AC O10282;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 73.1 kDa protein precursor (ORF21).
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OPMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H.; Russell R.R.; Funk C.J.; Evans J.; Harwood S.;
RA Rohrmann G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedrosis virus genome."
RL Virology 229:381-399(1997).
RN [2]
RP REVISIONS TO C-TERMINUS.
RA Rohrmann G.F.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U75930; AAC59020.2; -
KW Hypothetical protein; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 657 HYPOTHETICAL 73.1 KDA PROTEIN.
SQ SEQUENCE 657 AA; 73128 MW; 3B81055F4C27F8E4 CRC64;

Query Match 77.6%; Score 38; DB 1; Length 657;
Best Local Similarity 87.5%; Pred. No. 8.9;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PIDDGKGL 9
|||||
Db 25 PIDDGAGL 32

RESULT 4
ATPB_LACCA

ID ATPB_LACCA STANDARD; PRT; 470 AA.
AC Q03234;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ATP synthase beta chain (EC 3.6.3.14).
GN ATPD.
OS Lactobacillus casei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1582;
RN [1]
RP SEQUENCE FROM N.A.
RA Klugbauer N.; Ludwig W.; Bauerlein E.; Schleifer K.H.;
RT "Subunit beta of adenosine triphosphate synthase of Pectinatus frisingensis and Lactobacillus casei."
RL Syst. Appl. Microbiol. 15:323-330(1992).
CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC SUBUNIT.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(-)(In) = ADP + phosphate + H(+)(Out).
CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC

allele of the parE gene encoding a subunit of topoisomerase IV. ";
 Antimicrob. Agents Chemother. 41:173-179(1997).
 -!- FUNCTION: TOPOISOMERASE IV IS ESSENTIAL FOR CHROMOSOME
 SEGREGATION. IT HAS RELAXATION OF SUPERCOILED DNA ACTIVITY.
 PERFORMS THE DECATENATION EVENTS REQUIRED DURING THE REPLICATION
 OF A CIRCULAR DNA MOLECULE.
 -!- SUBUNIT: COMPOSED OF TWO SUBUNITS: PARC AND PARE.
 -!- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; M58409; AAA24298.1; ALT_SEQ.
 EMBL; U28377; AAA69198.1; .
 EMBL; AE000385; AAC76066.1; .
 EMBL; L22026; AAC36841.1; .
 PIR; B36075; B36075.
 HSP; P06982; 1A36.
 EcoGene; EG10687; parE.
 InterPro; IPR002288; DNA_gyraseB_C.
 InterPro; IPR001241; DNA_topoisol.
 Pfam; PF00986; DNA_gyraseB_C; 1.
 Pfam; PF00204; DNA_topoisolII; 1.
 Pfam; PF02518; HATPase_C; 1.
 PRINTS; PR00418; TP12FAMILY.
 ProDom; PD000616; DNA_topoisolII; 1.
 ProDom; PD149633; DNA_gyraseB_C; 1.
 SMART; SM00387; HATPase_C; 1.
 SMART; SM00433; TOP2c; 1.
 PROSITE; PS00177; TOPOISOMERASE_II; 1.
 Isomerase; Topoisomerase; ATP-binding; Antibiotic resistance;
 Complete proteome. 445
 VARIANT 445 L-> H (IN QUINOLONE-RESISTANT STRAIN
 DH161).
 SEQUENCE 630 AA: 70243 MW: 7F83D108BC1C6A41 CRC64;

 Query Match 71.4%; Score 35; DB 1; Length 630;
 Best Local Similarity 66.7%; Pred. No. 33;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 1 EPIDCKGL 9
 | | | | |
 66 EVIDDGRM 74

 RESULT 7
 ARE_SALTY STANDARD; PRT; 630 AA.
 P31598;
 01-JUL-1993 (Rel. 26, Created)
 01-MAR-2002 (Rel. 41, Last sequence update)
 01-MAR-2002 (Rel. 41, Last annotation update)
 Topoisomerase IV subunit B (EC 5.99.1.1).
 PARE OR STM3181 OR STV3359.
 Salmonella typhimurium, and
 Salmonella typhi.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Salmonella.
 NCBI_TaxID=602, 601;
 [1]

 SEQUENCE FROM N.A.
 SPECIES=S. typhimurium; STRAIN=LPT2;
 MEDLINE=94261812; PubMed=8388096;
 Springer A.L., Schmid M.B.;
 "Molecular characterization of the Salmonella typhimurium parE gene.";
 Nucleic Acids Res. 21:1805-1809(1993).

RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. typhimurium; STRAIN=LPT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.";
 RT Nature 413:852-856(2001).
 RL [3]
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. typhi; STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
 Baker S., Basham D., Brooks R.M., Chillingworth T., Connerton P.,
 Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 Whitehead S., Barrall B.G.;
 RA "Complete genome sequence of a multiple drug resistant Salmonella
 enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 CC -!- FUNCTION: TOPOISOMERASE IV IS ESSENTIAL FOR CHROMOSOME
 SEGREGATION. IT HAS RELAXATION OF SUPERCOILED DNA ACTIVITY.
 PERFORMS THE DECATENATION EVENTS REQUIRED DURING THE REPLICATION
 OF A CIRCULAR DNA MOLECULE.
 CC -!- SUBUNIT: COMPOSED OF TWO SUBUNITS: PARC AND PARE.
 CC -!- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; L05544; AAA27182.1; .
 EMBL; AE008846; AAL22055.1; .
 EMBL; AL627277; CAD03013.1; .
 PIR; S33711; S33711.
 HSP; P06982; 1A36.
 StyGene; SG10277; parE.
 InterPro; IPR002288; DNA_gyraseB_C.
 InterPro; IPR001241; DNA_topoisol.
 InterPro; IPR003594; HATPase_C.
 Pfam; PF00986; DNA_gyraseB_C; 1.
 Pfam; PF0204; DNA_topoisol; 1.
 Pfam; PF02518; HATPase_C; 1.
 PRINTS; PR00418; TP12FAMILY.
 ProDom; PD000616; DNA_topoisolII; 1.
 ProDom; PD149633; DNA_gyraseB_C; 1.
 SMART; SM00387; HATPase_C; 1.
 SMART; SM00433; TOP2c; 1.
 PROSITE; PS00177; TOPOISOMERASE_II; 1.
 Isomerase; Topoisomerase; ATP-binding; Complete proteome.
 VARIANT 67 67 V-> M (IN TS MUTANTS PARE206 & PARE374).
 VARIANT 399 399 G-> S (IN TS MUTANT PARE377).
 VARIANT 583 583 T-> P (IN TS MUTANT PARE493).
 CONFLICT 305 305 T-> Q (IN REF. 1).
 CONFLICT 333 333 T-> P (IN REF. 1).
 CONFLICT 383 383 A-> R (IN REF. 1).
 SEQUENCE 630 AA: 70088 MW: 7F2FCC25EB930019 CRC64;

Query Match 71.4%; Score 35; DB 1; Length 630;
 Best Local Similarity 66.7%; Pred. No. 33;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

DE 1 EPIDGKGL 9
DE 1 111111
GN 66 EVIDGKGL 74
OS
OC
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN 11
RN SEQUENCE FROM N.A.
RP STRAIN=74-OR23-1A;
RC MEDLINE=91170267; PubMed=1826006;
RX Caratoli A., Romano N., Ballario P., Morelli G., Macino G.;
RA "The Neurospora crassa carotenoid biosynthetic gene (albino 3)
RT reveals highly conserved regions among prenilyltransferases.";
RL J. Biol. Chem. 266:5854-5859(1991).
CC -1- FUNCTION: CATALYZES THE TRANS-ADDITION OF THE THREE MOLECULES OF
CC IPP ONTO DNAPP TO FORM GERANYLGERANYL PYROPHOSPHATE.
CC -1- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl
CC diphosphate = diphosphate + geranyl diphosphate.
CC -1- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
CC = diphosphate + trans,trans-farnesyl diphosphate.
CC -1- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl
CC diphosphate = diphosphate + geranylgeranyl diphosphate.
CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- INDUCTION: BY BLUE LIGHT.
CC -1- SIMILARITY: BELONGS TO THE PPP/GGPP SYNTHETASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U20940; AAC13867.1;
CC FIR: S15662; S15662.
CC InterPro: IPR000092; polyprenyl_synt.
CC Pfam: PF00348; polyprenyl_synt_1.
CC PROSITE: PS00444; POLYPRENYL_SYNTHET_2; 1.
CC PROSITE: PS00723; POLYPRENYL_SYNTHET_1; 1.
CC Carotenoid biosynthesis; isoprene biosynthesis; Transferase.
FT ACT_SITE 268 268 BY SIMILARITY.
SQ SEQUENCE 428 AA: 47887 MW; 7989DEEABD8E360F CRC64;

Query Match 69.4%; Score 34; DB 1; Length 428;
Best Local Similarity 75.0%; Pred. No. 35;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EPIDGKGL 8
DB 393 EDIDGKGL 400

RESULT 10
GDIA_BOVIN
ID GDIA_BOVIN STANDARD; PRT; 447 AA.
AC P21856;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE RAB GDP dissociation inhibitor alpha (RAB GDI alpha) (SMG P25A
DE GDI).
GN GD11 OR RABGDI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN 11
RN SEQUENCE FROM N.A.
RP

```

```

MEDLINE=90318376; PubMed=21151118;
Matsui Y., Kikuchi A., Araki S., Hata Y., Kondo J., Teranishi Y.,
Takai Y.;
" Moleculer cloning and characterization of a novel type of regulatory
protein (GDI1) for smg p25A, a ras p21-like GTP-binding protein.";
Mol. Cell. Biol. 10:4116-4122(1990).
[2]
SIMILARITY TO CHOROIDEAEMIA PROTEIN.
MEDLINE=91270365; PubMed=1904992;
Fodor E., Lee R.T., O'Donnell J.J.;
"Analysis of choroideraemia gene.";
Nature 351:614-614(1991).
[3]
X-RAY CRYSTALLOGRAPHY (1.81 ANGSTROMS).
MEDLINE=96196507; PubMed=8609986;
Schalk I., Zeng K., Wu S.-K., Stura E.A., Matteson J., Huang M.,
Tandon A., Wilson I.A., Balch W.E.;
"Structure and mutational analysis of Rab GDP-dissociation
inhibitor.";
Nature 381:42-48(1996).
-!- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
SUBSEQUENT BINDING OF GTP TO THEM.
-!- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
INHIBITOR.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL: D90103; BAA14134.1; .
PIR: A35652; A35652
PDB: 1GND; 12-FEB-97.
InterPro: IPR002005; Rab_GDI_REP.
Pfam: PF00996; GDI; 1.
PRINTS: PR00891; RABGDIREP.
GTPase activation; 3D-structure.
SEQUENCE 447 AA; 50565 MW; 59C68B03BDC1BA3F CRC64;

Query Match 69.4%; Score 34; DB 1; Length 447;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 EPIDDG 6
|||||
388 EPIDDG 393

SULT 11
IA_CANFA STANDARD; PRT; 447 AA.
GDI1;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
RAB GDP dissociation inhibitor alpha (RAB GDI alpha) (GDI-1).
GDI1.
Canis familiaris (Dog).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
NCBI_TaxID=9615;
[1]
SEQUENCE FROM N.A.
MEDLINE=99019719; PubMed=9802909;
Chen W., Feng Y., Chen D., Wandinger-Ness A.;
"Rab11 is required for trans-golgi network-to-plasma membrane
transport and a preferential target for GDP dissociation inhibitor.";
Mol. Biol. Cell 9:3241-3257(1998).

```

```

-!- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
SUBSEQUENT BINDING OF GTP TO THEM.
-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
-!- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
INHIBITOR.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL: AF027360; AAD04246.1; .
HSSP: P21856; 1GND.
DR InterPro: IPR002005; Rab_GDI_REP.
DR Pfam: PF00996; GDI; 1.
DR PRINTS: PR00891; RABGDIREP.
KW GTPase activation.
SQ SEQUENCE 447 AA; 50520 MW; 10280DAD33E4BCD0 CRC64;

Query Match 69.4%; Score 34; DB 1; Length 447;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIDDG 6
|||||
DB 388 EPIDDG 393

RESULT 12
GDI1_HUMAN STANDARD; PRT; 447 AA.
ID GDI1_HUMAN
AC P31150; P50394;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RAB GDP dissociation inhibitor alpha (RAB GDI alpha) (GDI-1) (XAP-4).
CN GDI1 OR RABGDI1 OR XAP4.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95152170; PubMed=7849400;
RA Sedlacek Z., Konecki D.S., Korn B., Klauck S.M., Poustka A.;
"Evolutionary conservation and genomic organization of XAP-4, an Xq28
located gene coding for a human rab GDP-dissociation inhibitor
(GDI).";
RL Mamm. Genome 5:633-639(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE-Relina;
RA MEDLINE=96062207; PubMed=7585614;
RA Nishimura N., Goji J., Nakamura H., Orita S., Takai Y., Sano K.;
"Cloning of a brain-type isoform of human Rab GDI and its expression
in human neuroblastoma cell lines and tumor specimens.";
RL Cancer Res. 55:5445-5450(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96311563; PubMed=8733135;
RA Chen E.Y., Zollo M., Mazzarella R.A., Ciccodicola A., Chen C.-N.,
Zuo L., Heiner C., Burrough F.W., Ripetto M., Schlessinger D.,
D'Urso M.;
"Long-range sequence analysis in Xq28: thirteen known and six
candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and
G6PD loci.";
RL Hum. Mol. Genet. 5:659-668(1996).
RN [4]

```

IP SEQUENCE FROM N.A.
 IC TISSUE=Lung;
 IC Strausberg R.;
 IC Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 IC [5]
 IC SEQUENCE OF 143-181 FROM N.A.
 IC Hochgeschwender U.;
 IC Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 IC [6]
 IC SEQUENCE OF 328-436 FROM N.A.
 IC Bhat K.S.;
 IC Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
 IC [7]
 IC SEQUENCE OF 349-361.
 IC TISSUE=Keratinocytes;
 IC MEDLINE=93162043; PubMed=1286667;
 IC Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
 IC Vanderkroghve J.;
 IC "Microsequences of 145 proteins recorded in the two-dimensional gel
 IC protein database of normal human epidermal keratinocytes";
 IC Electrophoresis 13:960-969(1992).
 IC [8]
 IC TISSUE SPECIFICITY.
 IC MEDLINE=95359978; PubMed=7543319;
 IC Bachner D., Sedlacek Z., Korn R., Hameister H., Poustka A.;
 IC "Expression patterns of two human genes coding for different rab GDP-
 IC dissociation inhibitors (GDIS), extremely conserved proteins involved
 IC in cellular transport";
 IC Hum. Mol. Genet. 4:701-708(1995).
 IC [9]
 IC VARIANT XLMR PRO-92.
 IC MEDLINE=98282090; PubMed=9620768;
 IC D'Adamo P., Menegon A., Lo Nigro C., Grassi M., Gullisano M.,
 IC Tamanini F., Bienvenu T., Gedeon A.K., Costra B., Wu S.-K., Tandon A.,
 IC Vaitorta F., Balch W.E., Chelly J., Toniolo D.;
 IC "Mutations in GDIL are responsible for X-linked non-specific mental
 IC retardation";
 IC Nat. Genet. 19:134-139(1998).
 IC [10]
 IC VARIANT XLMR PRO-423.
 IC MEDLINE=98334551; PubMed=9668174;
 IC Bienvenu T., Des Portes V., Saint Martin A., McDonnell N., Billuart P.,
 IC Carrie A., Vinet M.-C., Couvert P., Toniolo D., Rogers H.-H.,
 IC Moraine C., van Bokhoven H., Frys J.-P., Kahn A., Beidjord C.,
 IC Chelly J.;
 IC "Non-specific X-linked semidominant mental retardation by mutations in
 IC a Rab GDP-dissociation inhibitor";
 IC Hum. Mol. Genet. 7:1311-1315(1998).
 IC -I- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
 IC PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
 IC SUBSEQUENT BINDING OF GTP TO THEM.
 IC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 IC -I- TISSUE SPECIFICITY: BRAIN; PREDOMINANT IN NEURAL AND SENSORY
 IC TISSUES.
 IC -I- DISEASE: DEFECTS IN GDIL ARE THE CAUSE OF X-LINKED NON-SPECIFIC
 IC MENTAL RETARDATION (XLMR).
 IC -I- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
 IC INHIBITOR.
 IC -----
 IC This SWISS-PROT entry is copyright. It is produced through a collaboration
 IC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 IC the European Bioinformatics Institute. There are no restrictions on its
 IC use by non-profit institutions as long as its content is in no way
 IC modified and this statement is not removed. Usage by and for commercial
 IC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 IC or send an email to license@isb-sib.ch).
 IC -----
 IC EMBL: X79354; CAA55909.1; JOINED.
 IC EMBL: X79355; CAA55909.1; JOINED.
 IC EMBL: X79356; CAA55909.1; JOINED.
 IC EMBL: X79357; CAA55909.1; JOINED.
 IC EMBL: X79358; CAA55909.1; JOINED.
 IC EMBL: X79359; CAA55909.1; JOINED.

DR EMBL: X79360; CAA55909.1; JOINED.
 DR EMBL: X79364; CAA55909.1; JOINED.
 DR EMBL: X79361; CAA55909.1; JOINED.
 DR EMBL: X79362; CAA55909.1; JOINED.
 DR EMBL: X79363; CAA55909.1; JOINED.
 DR EMBL: X79353; CAA55908.1; -.
 DR EMBL: L4140; AAA92648.1; -.
 DR EMBL: BC000317; AA00317.1; -.
 DR EMBL: U14623; AAA21558.1; -.
 DR EMBL: D45021; BAA08078.1; -.
 DR EMBL: L05086; AAC15851.1; -.
 DR HSSP: P21856; JGND.
 DR Aarhus/Ghent-2DPAGE; 8408; IEF.
 DR MIM: 300104; -.
 DR MIM: 309541; -.
 DR InterPro: IPR002005; Rab_GDI_REP.
 DR Pfam: PF00996; GDI; 1.
 DR PRINTS: PRO0891; RABGDIREF.
 DR KW GTPase activation; Disease mutation.
 DR VARIANT 92 92
 FT L -> P (IN XLMR; CAUSES REDUCED BINDING
 FT AND RECYCLING OF RAB3A).
 FT /FTID-VAR_008130.
 FT R -> P (IN XLMR).
 FT /FTID-VAR_008131.
 FT D -> G (IN REF. 2).
 FT N -> K (IN REF. 2).
 FT NFD -> GTV (IN REF. 2).
 FT G -> V (IN REF. 5).
 FT H -> Q (IN REF. 5).
 FT D -> G (IN REF. 5).
 FT F -> S (IN REF. 2).
 SQ SEQUENCE 447 AA; 50582 MW; BC283A445E50A652 CRC64;
 Query Match 59.4%; Score 34; DB 1; Length 447;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EPIDDG 6
 DB 388 EPIDDG 393
 RESULT 13
 ID GDIA_RAT STANDARD; PRT; 447 AA.
 AC P50398; Q9R274;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE RAB GDP dissociation inhibitor alpha (RAB GDI alpha) (GDI-1).
 GN GD11 OK RABGDI.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
 RX MEDLINE=94245743; PubMed=8188702;
 RA Nishimura N., Nakamura H., Takai Y., Sano K.;
 RT "Molecular cloning and characterization of two rab GDI species from
 RT rat brain: brain-specific and ubiquitous types.";
 RL J. Biol. Chem. 269:14191-14198(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=94217740; PubMed=7513052;
 RA Shisheva A., Suedhof T.C., Czech M.P.;
 RT "Cloning, characterization, and expression of a novel GDP
 RT dissociation inhibitor isoform from skeletal muscle.";
 RL Mol. Cell. Biol. 14:3459-3468(1994).
 RN [3]

SEQUENCE OF 81-439 FROM N.A.
 TISSUE-Pancreas; PubMed-10996854;
 MEDLINE-20453283; PubMed-10996854;
 Caillol N., Pasqualini E., Lioubes R., Lombardo D.;
 "Impairment of bile salt-dependent lipase secretion in human
 pancreatic tumoral SOJ-6 cells."
 J. Cell. Biochem. 79:528-647(2000).
 -!- FUNCTION: REGULATES THE GTP/GTP EXCHANGE REACTION OF MOST RAB
 PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
 SUBSEQUENT BINDING OF GTP TO THEM.
 -!- SUBCELLULAR LOCATION: Cytoplasmic.
 -!- TISSUE SPECIFICITY: HIGH EXPRESSION IN BRAIN, LOWER IN OTHER
 TISSUES.
 -!- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
 INHIBITOR.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL: X74402; CAA52413.1; .
 EMBL: U07952; AAB16909.1; .
 EMBL: AF130987; AAD25536.1; .
 HSSP: P21856; LGND.
 InterPro: IPR002005; Rab_GDI_REP.
 Pfam: PF00996; GDI; 1.
 PRINTS: P00891; RABGDIREP.
 GTPase activation. 199 D -> Y (IN REF. 1).
 CONFLICT 230 230 G -> S (IN REF. 2).
 CONFLICT 330 330 S -> P (IN REF. 2).
 SEQUENCE 447 AA; 50536 MW; 58384671991DF793 CRC64;

Query Match 69.4%; Score 34; DB 1; Length 447;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 / 1 EPIDDC 6
 / 388 EPIDDC 393
 / |||||

RESULT 14
 K06_HUMAN
 Q1659; STANDARD; PRT; 721 AA.
 01-NOV-1997 (Rel. 35, Created)
 01-NOV-1997 (Rel. 35, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 Mitogen-activated protein kinase 6 (EC 2.7.1.1-) (Extracellular signal-
 regulated kinase 3) (ERK-3) (MAP kinase isoform p97) (p97-MAPK).
 MAPK6 OR PRK6 OR ERK3.
 Homo sapiens (Human).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 TISSUE-Fetal skeletal muscle;
 MEDLINE-95059049; PubMed-7969157;
 Zhu A.X., Zhao Y., Moller D.E., Flier J.S.;
 "Cloning and characterization of p97MAPK, a novel human homolog of
 rat ERK-3.";
 Mol. Cell. Biol. 14:8202-8211(1994).
 [2]
 SEQUENCE FROM N.A.
 TISSUE-Smooth muscle;
 MEDLINE-97030046; PubMed-8875998;

RA Meloche S., Beatty B.G., Pellerin J.;
 RT "Primary structure, expression and chromosomal locus of a human
 homolog of rat ERK3.";
 RL Oncogene 13:1575-1579(1996).
 CC -!- FUNCTION: PHOSPHORYLATES MICROTUBULE-ASSOCIATED PROTEIN-2 (MAP2).
 CC MAY PROMOTE ENTRY IN THE CELL CYCLE (BY SIMILARITY).
 CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN THE SKELETAL MUSCLE,
 CC FOLLOWED BY THE BRAIN. ALSO FOUND IN HEART, PLACENTA, LUNG, LIVER,
 CC PANCREAS, KIDNEY, AND SKIN FIBROBLASTS.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MAP KINASE SUBFAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL: X80692; CAA56709.1; .
 EMBL: U77964; AAA98769.1; .
 HSSP: P27703; IERK.
 DR MIM: 602904;
 DR InterPro: IPR000719; Euk_pkinase
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR Pfam: PF00669; pkinase; 1.
 DR SMART: SM00220; S_TKc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding; Cell cycle.
 FT DOMAIN 20 316 PROTEIN KINASE.
 FT NP_BIND 26 34 ATP (BY SIMILARITY).
 FT BINDING 49 49 ATP (BY SIMILARITY).
 FT ACT_SITE 152 152 BY SIMILARITY.
 SQ SEQUENCE 721 AA; 82680 MW; DAA3AAA9B98B831F CRC64;

Query Match 69.4%; Score 34; DB 1; Length 721;
 Best Local Similarity 71.4%; Pred. No. 60;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Caps 0;
 QY 1 EPIDDC 7
 QY ||:||||
 DB 645 EPVEDCK 651

RESULT 15
 RL28_HUMAN
 ID RL28_HUMAN STANDARD; PRT; 136 AA.
 AC P46779;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 60S ribosomal protein L28.
 GN RPL28.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Colon;
 RX MEDLINE-95290496; PubMed-7772601;
 RA Frigerio J.-M., Dagorn J.C., Iovanna J.L.;
 RT "Cloning, sequencing and expression of the L5, L21, L27a, L28, S5,
 RT S9, S10 and S29 human ribosomal protein mRNAs.";
 RL Biochim. Biophys. Acta 1262:64-68(1995).
 CC -!- SIMILARITY: BELONGS TO THE L28E FAMILY OF RIBOSOMAL PROTEINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 R EMBL; U14969; AAA85657.1; -
 R MIM; 603638; -
 R InterPro; IPR002672; Ribosomal_L28e.
 R Pfam; PF01778; Ribosomal_L28e; 1.
 R ProDom; PD010767; Ribosomal_L28e; 1.
 W Ribosomal protein... 0 BY SIMILARITY.
 T INIT_MET 0
 Q SEQUENCE 136 AA; 15630 MW; 6551B12C473D2342 CRC64;

Query Match 67.3%; Score 33; DB 1; Length 136;
 Best Local Similarity 66.7%; Pred. No. 16;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

y 1 EPIDDKGKL 9
 || ||||
 b 51 EPAADGKGV 59

Search completed: August 6, 2002, 17:07:36
 Job time: 878 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

M protein - protein search, using sw model

un on: August. 6, 2002, 17:09:30 ; Search time 111.35 Seconds

(without alignments)
13.983 Million cell updates/sec

itle: US-10-020-139-2_COPY_115_123

effect score: 49

equences: 1 EPIDGKGL 9

coring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

earched: 562222 seqs, 172994929 residues

otal number of hits satisfying chosen parameters: 562222

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

SPREMBL_19.*

- 1: sp-archaea.*
- 2: sp-bacteria.*
- 3: sp-fungi.*
- 4: sp-human.*
- 5: sp-invertebrate.*
- 6: sp-mammal.*
- 7: sp-mhc.*
- 8: sp-organelle.*
- 9: sp-phage.*
- 10: sp-plant.*
- 11: sp-rodent.*
- 12: sp-virus.*
- 13: sp-vertebrate.*
- 14: sp-unclassified.*
- 15: sp-rvirus.*
- 16: sp-bacteriap.*
- 17: sp-archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	49	100.0	249	Q9BQ00	Q9BQ00 homo sapien
2	49	100.0	249	Q96DR5	Q96DR5 homo sapien
3	42	85.7	424	Q30968	Q30968 nostoc punc
4	38	77.6	152	Q33C08	Q33C08 staphylococ
5	38	77.6	176	Q9N102	Q9N102 drosophila
6	38	77.6	299	Q9W5R8	Q9W5R8 drosophila
7	38	77.6	2167	Q9V3Z6	Q9V3Z6 drosophila
8	37	75.5	391	Q9SFE7	Q9SFE7 arabidopsis
9	37	75.5	766	Q9A2L2	Q9A2L2 caulobacter
10	36	73.5	410	Q926D3	Q926D3 rhizobium m
11	36	73.5	456	Q9H114	Q9H114 pseudomonas
12	36	73.5	744	Q69995	Q69995 streptomyce
13	36	73.5	1421	Q9FZU3	Q9FZU3 neisseria m
14	36	73.5	5627	Q9I120	Q9I120 pseudomonas
15	35	71.4	113	Q38236	Q38236 human immun
16	35	71.4	113	Q38238	Q38238 human immun

17	35	71.4	154	16	Q98H58	Q98H58 rhizobium l
18	35	71.4	258	15	P88400	P88400 human immun
19	35	71.4	387	16	Q45606	Q45606 bacillus su
20	35	71.4	614	5	Q20040	Q20040 caenorhabdi
21	35	71.4	626	16	Q9KPD9	Q9KPD9 vibrio chol
22	35	71.4	629	2	Q9ZNP9	Q9ZNP9 pseudomonas
23	35	71.4	629	16	Q9HUJ8	Q9HUJ8 pseudomonas
24	35	71.4	758	10	Q9SLW0	Q9SLW0 taraxacum o
25	35	71.4	2910	5	Q26008	Q26008 plasmodium
26	34	69.4	75	15	Q38232	Q38232 human immun
27	34	69.4	93	11	Q91Z41	Q91Z41 mus musculu
28	34	69.4	98	15	Q38227	Q38227 human immun
29	34	69.4	98	15	Q38231	Q38231 human immun
30	34	69.4	98	15	Q38234	Q38234 human immun
31	34	69.4	113	4	Q9UJ34	Q9UJ34 homo sapien
32	34	69.4	113	15	Q38242	Q38242 human immun
33	34	69.4	113	15	Q38244	Q38244 human immun
34	34	69.4	132	17	Q58790	Q58790 methanococc
35	34	69.4	282	2	Q9KGG2	Q9KGG2 lactobacill
36	34	69.4	282	16	Q99TB4	Q99TB4 staphylococ
37	34	69.4	291	2	Q9ADQ4	Q9ADQ4 streptomyce
38	34	69.4	313	12	Q40940	Q40940 kaposi's sa
39	34	69.4	367	16	Q9KD94	Q9KD94 bacillus ha
40	34	69.4	443	5	Q9VLB7	Q9VLB7 drosophila
41	34	69.4	447	4	Q96CX5	Q96CX5 homo sapien
42	34	69.4	447	11	Q91Y71	Q91Y71 mus musculu
43	34	69.4	448	5	Q24349	Q24349 drosophila
44	34	69.4	464	10	Q9MAP9	Q9MAP9 arabidopsis
45	34	69.4	465	5	Q9VHJ8	Q9VHJ8 drosophila

ALIGNMENTS

RESULT 1
Q9BQ00 PRELIMINARY; PRT; 249 AA.
AC Q9BQ00;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BA49G10.1 (SIMILAR TO BOVINE SALIVARY PROTEIN BSP30).
GN BA49G10.1.
OS Homo sapiens (Human).
OC Fukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tracey A.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL121901; CAC03546.1; .
SQ SEQUENCE 249 AA; 27011 MW; E64E0794A1B4DB7D CRC64;

Query Match 100.0%; Score 49; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EPIDGKGL 9
Db 115 EPIDGKGL 123
|||||

RESULT 2 PRELIMINARY; PRT; 249 AA.
Q96DR5
ID Q96DR5
AC Q96DR5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PAROTID SECRETORY PROTEIN.
OS Homo sapiens (Human).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 TISSUE=PAROTID;
 Venkatesh S.G., Geetha C., Gorr S.-U.,
 "A member of the PSP/plunc family of BP1 proteins is expressed in the
 human parotid gland."
 Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 EMBL: AF432917; AAL28113.1;
 SEQUENCE 249 AA; 27110 MW; FD54B624A1A4CA7C CRC64;

Query Match 100.0%; Score 49; DB 4; Length 249;
 Best Local Similarity 100.0%; Pred. No. 0.28;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 EPIDDGKGL 9
 |||||

115 EPIDDGKGL 123

SULT 3

030968 PRELIMINARY; PRT; 424 AA.
 030968;

01-JAN-1998 (TREMBLrel. 05, Created)

01-JAN-1998 (TREMBLrel. 05, Last sequence update)

01-DEC-2001 (TREMBLrel. 19, Last annotation update)

CARBOXYL TERMINAL PROTEASE.

CTPH.

Nostoc punctiforme.

Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.

NCBI_TaxID=63737;

[1]

SEQUENCE FROM N.A.

STRAIN=ATCC29133;

MEDLINE=98406048; PubMed=9733698;

Campbell E.L., Brahamsha B., Weeks J.C.;

"Mutation of an alternative sigma factor in the cyanobacterium nostoc
 punctiforme results in increased infection of its symbiotic plant
 partner, anthoceros punctatus."

J. Bacteriol. 180:4938-4941(1998).

EMBL: AF022823; AAC45366.1;

MEROPS; S41.002;

InterPro: IPR001478; PDZ.

InterPro: IPR003581; TSPC.

Pfam: PF00595; PDZ; 1

SMART: SM00228; PDZ; 1

SMART: SM00245; TSPC; 1.

PROSITE: PS50106; PDZ; 1.

Protease.

SEQUENCE 424 AA; 45830 MW; C1736347D3EFCD1 CRC64;

Query Match 85.7%; Score 42; DB 2; Length 424;
 Best Local Similarity 77.8%; Pred. No. 10;
 Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

1 EPIDDGKGL 9

|||||

344 EPLDDGKGL 352

SULT 4

3CC8

Q93CC8 PRELIMINARY; PRT; 152 AA.

Q93CC8;

01-DEC-2001 (TREMBLrel. 19, Created)

01-DEC-2001 (TREMBLrel. 19, Last sequence update)

01-DEC-2001 (TREMBLrel. 19, Last annotation update)

HYPOTHETICAL 18.1 KDA PROTEIN.

OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 NCBI_TaxID=1280;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=COL;
 RA Yarwood J.M., McCormick J.K., Paustian M.L., Orwin P.M., Kapur V.,
 Schlievert P.M.;
 "Staphylococcus aureus pathogenicity island 3 (SaPI3).";
 Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 EMBL: AF410775; AAL04144.1;
 KW Hypothetical protein.
 SEQUENCE 152 AA; 18082 MW; 17A63018DE9B9AB3 CRC64;

Query Match 77.6%; Score 38; DB 2; Length 152;
 Best Local Similarity 87.5%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PIDDGKGL 9

|||||

DB 15 PIDDGKSL 22

RESULT 5

Q9NIU2

ID Q9NIU2 PRELIMINARY; PRT; 176 AA.

AC Q9NIU2;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE YIPPEE INTERACTING PROTEIN 6 (FRAGMENT).

GN YIP6 OR CG17489.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=IMAGINAL DISCS;

RA Roxstrom-Lindquist K., Faye I.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF195190; AAF27819.1;

DR FLYBase: FBgn040061; Yip6.

DR InterPro: IPR001149; Ribosomal_L18p.

DR PRINTS: PR00058; RIBOSOMAL15.

DR ProDom: PD001394; Ribosomal_L18p; 1.

FT NON_TER 1

FT NON_TER 176 176

SQ SEQUENCE 176 AA; 19791 MW; 73CE21AB75AE80A7 CRC64;

Query Match 77.6%; Score 38; DB 5; Length 176;
 Best Local Similarity 75.0%; Pred. No. 22;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EPIDDGKGL 8

|||||

DB 18 EPVDDGPG 25

RESULT 6

Q9W5R8

ID Q9W5R8 PRELIMINARY; PRT; 299 AA.

AC Q9W5R8;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE YIP6 PROTEIN.

GN YIP6 OR CG17489.

OS Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 CX NCBI_TaxID=7227;
 [1]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
 RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Mount S.M., Moy M., Murphy B., Murphry C., Morris J., Moshrefi A.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
 RA "The genome sequence of *Drosophila melanogaster*."
 RT Science 287:2185-2195(2000).
 RL
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-Y AND CN BW SP;
 RX MEDLINE=99403001; PubMed-10471707;
 RA Ashburner M., Misra S., Rote J., Root J., Lewis S.E., Blazej R., Davis T.,
 RA Doyle C., Galle R., George R.A., Harris N., Hartzell G., Harvey D.,
 RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
 RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
 RA Celniker S., Rubin G.M.:
 RA "An exploration of the sequence of a 2.9-Mb region of the genome of
 RT *Drosophila melanogaster*: the *Adh* region."
 RL Genetics 153:179-219(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Y AND CN BW SP;
 RA Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
 RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
 RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
 RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
 RA Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
 RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
 RA Sethi H., Snir E., Swirskas R.R., Wan K.H., Weinburg T., Zhang R.,
 RA Zieran L.L., Rubin G.M.:
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003646; AAF53435.1; -;
 DR EMBL; AE003412; AAF4915.1; -;

C Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 C Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 C Ephydroidea; Drosophilidae; Drosophila.
 X NCBI_TaxID=7227;
 [1]
 N
 P SEQUENCE FROM N.A.
 C STRAIN=BERKELEY;
 X MEDLINE=20196006; PubMed-10731132;
 A Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 A Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 A George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 A Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 A Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 A Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 A Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
 A Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 A Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 A Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 A Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 A Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 A de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 A Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 A Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 A Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 A Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 A Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 A Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 A Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 A Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 A Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 A Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 A Merkulov G., Milshina N.V., Mobarry B., Murphy L., Muzny D.M., Nelson D.L.,
 A Mount S.M., Moy M., Murphy B., Murphry C., Morris J., Moshrefi A.,
 A Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 A Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 A Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 A Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 A Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 A Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 A Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 A Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 A Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 A Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 A Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
 A "The genome sequence of *Drosophila melanogaster*."
 T Science 287:2185-2195(2000).
 L EMBL; AE002910; AAF45418.2; -;
 R FlyBase; FBgn0040061; v1p6.
 R InterPro; IPR001149; Ribosomal_L18p.
 R Pfam; PF00861; Ribosomal_L18p; 1.
 R PRINTS; PR00058; RIBOSOMALL5.
 R PRODOM; PD001394; Ribosomal_L18p; 1.
 Q SEQUENCE 299 AA; 34037 MW; 523F50ED67E7841A CRC64;
 Query Match 77.6%; Score 38; DB 5; Length 299;
 Best Local Similarity 75.0%; Pred. No. 39;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Y 1 EPIDDKGK 8
 b 133 EPVDDGPG 140
 ESULT 7
 D QV326 PRELIMINARY; PRT; 2167 AA.
 C QV326;
 T 01-MAY-2000 (TrEMBLrel. 13, Created)
 T 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 T 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 E CK GENE PRODUCT.
 N CK OR BG:DS00929.11 OR CG7595.

```

HSSP: P08799; 1MND;
Flybase: FBgn0000317; ck.
InterPro: IPR000299; Band_4.1.
InterPro: IPR000048; IQ.
InterPro: IPR001609; myosin_head.
InterPro: IPR000857; MYTH4.
InterPro: IPR000567; SBP_bac_1.
InterPro: IPR001452; SH3.
Pfam: PF00612; IQ; 4.
Pfam: PF00663; myosin_head; 1.
Pfam: PF00784; MYTH4; 2.
Pfam: PF00181; SH3; 1.
PRINTS: PR00193; MYOSINHEAVY.
ProDom: PD000355; myosin_head; 1.
SMART: SM00295; B41; 2.
SMART: SM00015; IQ; 3.
SMART: SM00242; MYSC; 1.
SMART: SM00139; MYTH4; 2.
SMART: SM00326; SH3; 1.
PROSITE: PS00557; BAND_41_3; 2.
PROSITE: PS00096; IQ; 1.
PROSITE: PS01037; SBP_BACTERIAL_1; UNKNOWN_1.
PROSITE: PS00002; SH3; 1.
SEQUENCE 2167 AA; 250307 MW; 3C57E34ADDD89A42 CRC64;

Query Match 77.6%; Score 38; DB 5; Length 2167;
Best Local Similarity 85.7%; Pred. No. 3.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 EPIDGK 7
|||||
926 EPVDGK 932

SULT 8
SFE7
Q9SFE7 PRELIMINARY; PRT; 391 AA.
Q9SFE7;
01-MAY-2000 (TrEMBLrel. 13, Created)
01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
T26f17.11.
Arabidopsis thaliana (Mouse-ear cross).
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
eurosids II; Brassicales; Brassicaceae; Arabidopsi.
NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
Kim C., Altafi H., Bei B., Chin C., Chlou J., Choi E., Conn L.,
Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
Ecker J.R.;
"Genomic sequence for Arabidopsis thaliana BAC T26f17 from chromosome
I.";
Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
Ecker J.R.;
Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
Kim C., Altafi H., Bei B., Chin C., Chlou J., Choi E., Conn L.,
Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
Toriumi M., Vaysberg M., Yu G., Federspiel N., Theologis A.,
Ecker J.;

```

```

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chlou J., Choi E., Lam B.,
RA Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC013482; AAF16542.1; -.
DR InterPro: IPR000620; DUF6.
DR Pfam: PF00892; DUF6; 2.
SQ SEQUENCE 391 AA; 41990 MW; EBB4B07E4DD7B522 CRC64;

Query Match 75.5%; Score 37; DB 10; Length 391;
Best Local Similarity 66.7%; Pred. No. 82;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EPIDGKGL 9
:|:| |
Db 359 KPVDGKGL 367

RESULT 9
Q9A2L2 PRELIMINARY; PRT; 766 AA.
ID Q9A2L2
AC Q9A2L2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NADP-DEPENDENT MALIC ENZYME.
GN CC3549.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
[1]
SEQUENCE FROM N.A.
RC STRAIN-ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
"Complete genome sequence of Caulobacter crescentus.";
proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
EMBL: AE006013; AAK25511.1; -.
TIGR: CC3549; -.
DR InterPro: IPR000515; BPD_transp.
DR InterPro: IPR001891; Malic_enzyme.
DR InterPro: IPR002505; PTA_PTB.
DR Pfam: PF00390; malic; 1.
DR Pfam: PF01515; PTA_PTB; 1.
DR PROSITE: PS00402; BPD_TRANS_P_INN_MEMBR; UNKNOWN_1.
DR PROSITE: PS00331; MALIC_ENZYMES; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 766 AA; 82539 MW; 7B8E11FEB06577A9 CRC64;

Query Match 75.5%; Score 37; DB 16; Length 766;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIDGKGL 8
:|:| |
Db 759 QPVDGEG 766

```

```

RESULT 10
926D3 D Q926D3 PRELIMINARY: PRT: 410 AA.
C Q926D3:
T T 01-DEC-2001 (TREMBLrel. 19, Created)
T T 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
T T 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
E PROBABLE N-ACETYLURAMOVYL-L-ALANINE AMIDASE AMIC PRECURSOR
E TRANSMEMBRANE PROTEIN (EC 3.5.1.28).
S Rhizobium meliloti (Sinorhizobium meliloti).
C Bacteria: Proteobacteria: alpha subdivision; Rhizobiaceae group;
C Rhizobiaceae; Sinorhizobium.
X NCBI_TaxID=382;
P [1]
N SEQUENCE FROM N.A.
C STRAIN=1021;
X MEDLINE=21368234; PubMed=11474104;
A Calhert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
A Barloy-Hubier F., Barnett M.J., Becker A., Boistard P., Bothe G.,
A Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
A Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
A Gloux S., Godrie T., Goffeau A., Golding B., Guzy J., Gurjal M.,
A Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
A Kahn M.L., Kaiman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
A Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelie D., Purnelle B.,
A Ramsperger U., Surzycki R., Thebault P., Vandenbol M., Yeh K.-C., Batut J.,
A Vorhoefer F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
T "The composite genome of the legume symbiont Sinorhizobium meliloti.";
L Science 293:668-672(2001).
W EMBL: AL591786; CAC45908.1;
W Hydrolase: Complete proteome.
Q SEQUENCE 410 AA; 44588 MW; CF55FC972194E8E8 CRC64;

Query Match 73.5%; Score 36; DB 16; Length 410;
Best Local Similarity 62.5%; Pred. No. 1.3e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Y 1 EPIDGKG 8
b 128 KPVEDGKG 135
:|||||

RESULT 11
9HY14 D Q9HY14 PRELIMINARY: PRT: 456 AA.
C Q9HY14:
T 01-MAR-2001 (TREMBLrel. 16, Created)
T 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
T 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
E HYPOTHETICAL PROTEIN PA3421.
N PA3421.
S Pseudomonas aeruginosa.
C Bacteria: Proteobacteria: gamma subdivision; Pseudomonadaceae;
C Pseudomonas.
X NCBI_TaxID=287;
P [1]
N SEQUENCE FROM N.A.
C STRAIN=ATCC 15692 / PA01;
X MEDLINE=20437337; PubMed=10984043;
A Stover C.K., Pham X.-Q., Erwin A.L., Mizoguchi S.D., Warriner P.,
A Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
A Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
A Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
A Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
A Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
T "Complete genome sequence of Pseudomonas aeruginosa PA01, an
T opportunistic pathogen.";
L Nature 406:959-964(2000).
L EMBL: AE004763; RAG06809.1;
W Hypothetical protein; Complete proteome.
Q SEQUENCE 456 AA; 50690 MW; BFE2F230EC94F736 CRC64;

Query Match 73.5%; Score 36; DB 2; Length 744;
Best Local Similarity 62.5%; Pred. No. 2.5e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIDGKG 8
Db 341 EPLEDGRG 348
|||||:|

RESULT 13
Q9FZU3 D Q9FZU3 PRELIMINARY: PRT: 1421 AA.
C Q9FZU3:
AT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ORF 4.
OS Neisseria meningitidis phage 2120.
OC Viruses
X NCBI_TaxID=132905;
P [1]
N SEQUENCE FROM N.A.
RC STRAIN=2120;
RX MEDLINE=21172874; PubMed=11274117;

```

Claus H., Stoevesandt J., Frosch M., Vogel U.:
 "Genetic isolation of meningococci of the electrophoretic type 37
 complex";
 J. Bacteriol. 183:2570-2575(2001).
 EMBL: AJ278707; CAC19023.1;
 InterPro: IPR003961; FN_III.
 SMART: SM00060; FN3; 1.
 SEQUENCE 1421 AA; 153413 MW; 7A17FD4AEB6FA9A6 CRC64;

Query Match 73.5%; Score 36; DB 9; Length 1421;
 Best Local Similarity 75.0%; Pred. No. 5.2e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 1 EPIDDKGK 8
 |||||
 317 EPVDDGFG 324

SULT 14
 I120 PRELIMINARY; PRT: 5627 AA.
 Q91120
 01-MAR-2001 (TREMRLrel. 16, Created)
 01-MAR-2001 (TREMRLrel. 16, Last sequence update)
 01-OCT-2001 (TREMRLrel. 18, Last annotation update)
 HYPOTHETICAL PROTEIN PA2462.
 PA2462.
 Pseudomonas aeruginosa.
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.
 NCBI_TaxID=287;
 [1]

SEQUENCE FROM N.A.
 STRAIN-ATCC 15692 / PA01;
 MEDLINE-20437337; PubMed-10984043;
 Stover C.K., Pham X.-O., Erwin A.L., Mizoguchi S.D., Warren P.,
 Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
 Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
 "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen";
 Nature 406:959-964(2000).
 EMBL: AE004673; AAG05850.1;
 InterPro: IPR01899; Gram_pos_anchor.
 PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 Hypothetical protein; Complete proteome.
 SEQUENCE 5627 AA; 573167 MW; 2D2A1154C99D3750 CRC64;

Query Match 73.5%; Score 36; DB 16; Length 5627;
 Best Local Similarity 87.5%; Pred. No. 2.3e+03;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 2 PIDDKGKL 9
 |||||
 5280 PIDDKKCL 5287

SULT 15
 8236 PRELIMINARY; PRT: 113 AA.
 O38236
 01-JAN-1998 (TREMRLrel. 05, Created)
 01-JAN-1998 (TREMRLrel. 05, Last sequence update)
 01-DEC-2001 (TREMRLrel. 19, Last annotation update)
 GAG PROTEIN (FRAGMENT).
 GAG.
 Human immunodeficiency virus type 1.
 Virus; Retroviridae; Lentivirus.
 NCBI_TaxID=11676;

RN SEQUENCE FROM N.A.
 RP STRAIN-PATIENT 5;
 RX MEDLINE-97404676; PubMed-9261388;
 RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
 RA Vasudevarachari M.B., Salzman N.P.;
 RT "Drug resistance during Indinavir therapy is caused by mutations in
 RL the protease gene and in its Gag substrate cleavage sites";
 J. Virol. 71:6662-6670(1997).
 DR EMBL: AF024133; AAB83009.1;
 DR HSSP: P05888; IAAF
 DR InterPro: IPR001878; Znf_CCHC.
 DR Pfam: PF00098; zf-CCHC; 2.
 DR PRINTS: PR00939; C2HCZNFINGER.
 DR SMART: SM00343; Znf_C2HC; 2.
 KW Zinc-finger.
 FT NON_TER 1 1
 FT NON_TER 113 113
 SQ SEQUENCE 113 AA; 12888 MW; 671A99ED2DD5D8B9 CRC64;
 Query Match 71.4%; Score 35; DB 15; Length 113;
 Best Local Similarity 77.8%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 EPIDUGKGL 9
 |||||
 Db 100 EPIDKDKGL 108
 Search completed: August 6, 2002, 17:09:32
 Job time: 934 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

M protein - protein search, using sw model

on: August 6, 2002, 16:52:53 ; Search time 138.55 seconds
(without alignments)
7.215 Million cell updates/sec

file: US-10-020-139-2_COPY_115_123

effect score: 49

sequence: 1 EPIDGKGL 9

coring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 747574 seqs, 111073796 residues

total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

1:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	49	100.0	249	19	AAW69221 Human parotid secr
2	49	100.0	249	19	AAW60682 Human parotid secr
3	49	100.0	249	21	AAW24069 Human PRO1025 prot
4	49	100.0	249	21	AAW25765 Human secreted pro
5	49	100.0	249	21	AAW75351 Human secreted pro
6	49	100.0	260	22	AAW25745 Human protein sequ
7	38	77.6	299	22	ABW66320 Drosophila melanog
8	38	77.6	299	22	ABW67420 Drosophila melanog
9	38	77.6	2167	22	ABW60369 Drosophila melanog
10	35	71.4	98	22	ABG24503 Novel human diagno
11	35	71.4	127	22	ABG29553 Novel human diagno

12	35	71.4	296	22	AAU04903 Micromonospora eve
13	35	71.4	414	22	AAU57477 Propionibacterium
14	35	71.4	481	22	ABG29555 Novel human diagno
15	35	71.4	511	22	AAU41020 Propionibacterium
16	34	69.4	282	20	AAW73375 S. aureus D-alanin
17	34	69.4	443	22	ABW60633 Drosophila melanog
18	34	69.4	447	12	AAK12388 smg p25A GDP Disso
19	34	69.4	465	22	ABW76222 Drosophila melanog
20	34	69.4	746	22	ABG12604 Novel human diagno
21	33	67.3	72	22	ABG25931 Novel human diagno
22	33	67.3	112	21	AAU04013 Human secreted pro
23	33	67.3	113	21	AAU01932 Human secreted pro
24	33	67.3	119	21	AAU04011 Human secreted pro
25	33	67.3	131	18	AAW89776 Staphylococcus aur
26	33	67.3	140	22	ABG04277 Novel human diagno
27	33	67.3	143	21	AAU01930 Human secreted pro
28	33	67.3	144	21	AAU04012 Human secreted pro
29	33	67.3	155	22	AAU66994 Propionibacterium
30	33	67.3	155	22	AAU55191 Propionibacterium
31	33	67.3	168	22	AAU22294 Propionibacterium
32	33	67.3	186	22	AAU16627 Human novel secret
33	33	67.3	194	22	AAU16214 Human novel secret
34	33	67.3	301	22	ABW71849 Drosophila melanog
35	33	67.3	350	22	ABG04977 Novel human diagno
36	33	67.3	382	21	AAW08638 A human apoptosis
37	33	67.3	376	20	AAU35577 C. pneumoniae prot
38	33	67.3	396	16	AAW76550 Thermostable alkali
39	33	67.3	397	17	AAW93147 Bacillus sp. xylan
40	33	67.3	440	22	ABW60664 Drosophila melanog
41	33	67.3	578	19	AAW40046 Biologically equiv
42	33	67.3	578	19	AAW40047 Biologically equiv
43	33	67.3	578	19	AAW40048 Biologically equiv
44	33	67.3	578	19	AAW40049 Biologically equiv
45	33	67.3	578	19	AAW40050 Biologically equiv

ALIGNMENTS

RESULT 1
AAW69221
ID AAW69221 standard; Protein: 249 AA.

XX	AAW69221;
XX	16-OCT-1998 (first entry)
XX	Human parotid secretory protein.
DE	Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
XX	non-immune defensive disorder; immune system disorder; cancer; human;
KW	therapy; diagnosis.
XX	Homo sapiens.
OS	Homo sapiens.
XX	Key
XX	Location/Qualifiers
FT	Peptide
FT	1..18
FT	/note= "signal peptide"
FT	19..249
FT	/note= "mature hPSP"
XX	W09828420-A1.
XX	02-JUL-1998.
XX	18-DEC-1997; 97WO-US23522.
XX	23-DEC-1996; 96US-0034429.
XX	(HUMA-) HUMAN GENOME SCI INC.
XX	Duan R, Ruben SM;

WPI: 1998-377651/32.
N-PSDB; AAV44759.

New nucleic acid encoding human parotid secretory protein or its fragments - useful for diagnosis and treatment of, e.g. digestive and endocrine disorders and for drug screening

Claim 16; Fig 1: 94pp; English.

This sequence is the human parotid secretory protein (hPSP) of the invention. The hPSP DNA is useful for chromosome identification and isolation of the corresponding genomic DNA. The DNA and protein can be used to detect abnormal levels of hPSP (in standard blotting, amplification or immuno assays), particularly for diagnosis of digestive, non-immune defensive, endocrine or immune system disorders. A particular application is diagnosis of cancers of the salivary gland, thymus and pancreas which are associated with high levels of hPSP. The protein is also useful as antifungal, antibacterial, antiparasitic and antiviral agents and may be expressed in vivo from the DNA. The protein, or cells expressing it, are used in screening tests to identify specific antagonists, e.g. antibodies (Ab), polypeptides and antisense nucleic acids, which are potentially useful for treating conditions associated with excessive hPSP production. Cells containing the DNA are used to express the recombinant protein and this can be used to raise Ab, useful for diagnosis, therapy, for affinity purification and to identify hPSP-binding proteins.

Sequence 249 AA;

Query Match 100.0%; Score 49; DB 19; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.084;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 EPIDDDGKGL 9
|||||
115 epiddgkgl 123

SULT 2

*60682

AAW60682 standard; Protein; 249 AA.

AAW60682;

18-SEP-1998 (first entry)

Human parotid secretory protein (HSPS).

Parotid secretory protein; human; cancer; autoimmune disease;
secretory tissue; gastrointestinal tissue; hPSP; Sjogren's syndrome;
Graves disease; thyroiditis; insulin-dependent diabetes; pancreatitis;
ulcerative colitis; Crohn's disease; atrophic gastritis.

Homo sapiens.

WO9821329-A1.

22-MAY-1998.

07-NOV-1997; 97WO-US20651.

14-NOV-1996; 96US-0749288.

(INCY-) INCYTE PHARM INC.

Bandman O, Goli SK;

WPI: 1998-297933/26.

N-PSDB; AAV37699.

New parotid secretory protein - useful for, e.g. treatment of cancer and auto-immune disease, particularly of secretory or

PT gastrointestinal tissues

XX Claim 1: Fig 1A-C: 65pp; English.

XX This represents a human parotid secretory protein (hPSP). Antagonists
CC that bind specifically to, and modulate activity of hPSP are used to
CC treat cancer and autoimmune diseases particularly of secretory or
CC gastrointestinal tissue, e.g. cancer of salivary gland, thyroid,
CC prostate, breast, gastrointestinal tract or pancreas, Sjogren's syndrome,
CC Graves disease, thyroiditis, insulin-dependent diabetes, pancreatitis,
CC ulcerative colitis, Crohn's disease and atrophic gastritis. Host cells
CC containing expression vectors comprising the hPSP nucleic acid are used
CC to produce recombinant hPSP which is used to generate antibodies and to
CC screen for its antagonists. Antibodies are useful directly as
CC antagonists, to transport drugs to hPSP-expressing cells, to detect cells
CC that express hPSP, to monitor patients being treated with hPSP, and for
CC purification of hPSP from natural sources. Expression of hPSP may
CC indicate cell proliferation. hPSP nucleic acid or its fragments are used
CC to detect hPSP-encoding sequences (optionally after amplification by PCR)
CC by hybridisation, particularly for diagnosis and monitoring of disease,
CC but also for mapping the chromosomal sequence.

XX Sequence 249 AA;

Query Match 100.0%; Score 49; DB 19; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.084;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIDDDGKGL 9

|||||

Db 115 epiddgkgl 123

RESULT 3

AAB24069

ID AAB24069 standard; Protein; 249 AA.

XX AAB24069;

XX 29-JAN-2001 (first entry)

XX Human PRO1025 protein sequence SEQ ID NO:38.

XX Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
KW proliferation; tumorigenesis; identification; cancer; cytostatic;
KW neurotropic; neuroprotective; antiinflammatory; immunosuppressive;
KW immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
KW neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
KW hypothalamic disorder; glandular disorder; macrophagal disorder;
KW epithelial disorder; stromal disorder; blastocoelec disorder;
KW inflammatory disorder; immunologic disorder.

XX Homo sapiens.

XX WO200053755-A2.

XX 14-SEP-2000.

XX 06-JAN-2000; 2000WO-US00376.

XX 08-MAR-1999; 99WO-US05028.

XX 02-JUN-1999; 99WO-US12252.

XX 23-JUN-1999; 99US-0141037.

XX 07-JUL-1999; 99US-0143048.

XX 26-JUL-1999; 99US-0145698.

XX 30-NOV-1999; 99WO-US28313.

XX 20-DEC-1999; 99WO-US30911.

XX 05-JAN-2000; 2000WO-US00219.

(GETH) GENENTECH INC.

XX Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;

```

>I Watanabe CK, Wood WI;
>X
>X API: 2000-572270/53.
>X N-PSDB: AAC58379.
>X
>X Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
>X treatment, diagnosis and prevention of cancer.
>X
>X Claim 61; Fig 26; 286pp; English.
>X
>X The present invention describes an isolated antibody that binds to
>X one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO355,
>X PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
>X PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
>X PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
>X PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
>X growth. The PRO polypeptides and nucleotides are useful in the
>X treatment, diagnosis and prevention of cancer. The antibodies and other
>X anti-tumour compounds maybe used to treat various conditions, including
>X those characterised by overexpression and/or activation of the amplified
>X PRO genes. Exemplary conditions or disorders to be treated with such
>X antibodies and other compounds include benign or malignant tumours
>X (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
>X colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
>X carcinomas, sarcomas, glioblastomas, and various disorders such as neuronal,
>X leukaemias and lymphoid malignancies, other disorders such as neuronal,
>X gliial, astrocytal, hypochlaemic and other glandular, macrophagal,
>X epithelial, stromal and blastocoele disorders, and inflammatory,
>X angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
>X primers and hybridisation probes used in the isolation of the human PRO
>X sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
>X PRO polynucleotide and protein sequences given in the exemplification of
>X the present invention.
>X
>X Sequence 249 AA;
>X
>X Query Match 100.0%; Score 49; DB 21; Length 249;
>X Best Local Similarity 100.0%; Pred. No. 0.084;
>X Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
>X
>X >Y 1 EPIDGKGL 9
>X | | | | | | | |
>X >B 115 epidgkgl 123
>X
>X RESULT 4
>X AAB25765
>X ID ID AAB25765 standard; Protein; 249 AA.
>X
>X AC AAC25765;
>X
>X XX 28-NOV-2000 (first entry)
>X
>X XX Human secreted protein SEQ ID #77.
>X
>X XX Human; secreted protein; forensic procedure; gene therapy;
>X chromosome mapping; cancer; autoimmune disease; cardiovascular disorder;
>X cystic fibrosis; hypothyroidism; immunological disorder; amyloidosis;
>X brain disorder; skeletal muscle disorder; eye disorder; obesity;
>X mitochondriocytopathy; diabetes; atherosclerosis; Alzheimer's disease;
>X neurodegenerative disorder; graft rejection; dementia; hyperlipidaemia;
>X septic shock; impotence.
>X
>X XX Homo sapiens.
>X
>X XX WO200037491-A2.
>X
>X PN 29-JUN-2000.
>X
>X PD
>X
>X XX 20-DEC-1999; 99WO-IB02058.
>X
>X XX 22-DEC-1998; 98US-0113686.

```

PI Dumas Milne Edwards J, Bougueleret L, Robert S;
 XX WPI; 2001-071487/08.
 DR
 XX 49 Secreted proteins and the nucleic acids encoding them, useful in
 PT gene therapy and for detecting similar sequences in samples -
 XX
 XX Claim 10; Page 281; 307pp; English.
 PS
 XX The present invention relates to 49 secreted proteins and the cDNAs
 CC encoding them. The protein and nucleic acids may be used in the
 CC prevention, treatment and diagnosis of diseases associated with
 CC inappropriate protein expression.
 XX
 XX Sequence 249 AA;
 SQ
 Query Match 100.0%; Score 49; DB 22; Length 249;
 Best Local Similarity 100.0%; Pred. No. 0.084;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 2Y 1 EPIDGKGL 9
 IIIIIII
 3B 115 epiddgkgl 123
 RESULT 6
 XAM25745
 ID AAM25745 standard; Protein: 260 AA.
 AC
 XX AAM25745;
 XX
 XX 16-OCT-2001 (first entry)
 XX
 XX Human protein sequence SEQ ID NO:1260.
 XX
 XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 XX anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive;
 XX antibacterial; endocrine; cardiant; central nervous system; virucide;
 XX anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 XX antiagregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
 XX dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 XX neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 XX immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 XX antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 XX cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 XX genetic disease; haematopoietic disorder; platelet disorder; asthma,
 XX thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 XX allergic rhinitis; diabetes; multiple sclerosis; depression;
 XX Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 XX neurological disorder.
 XX
 XX Homo sapiens.
 XX
 XX WO200153455-A2.
 XX
 XX 26-JUL-2001.
 XX
 XX 22-DEC-2000; 2000WO-US35017.
 XX
 XX 23-DEC-1999; 99US-0471275.
 XX
 XX 21-JAN-2000; 2000US-0488725.
 XX
 XX 25-APR-2000; 2000US-0552317.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT;
 XX
 XX WPI; 2001-457603/49.
 XX
 XX N-FSDB; AAH99686.
 XX
 XX Isolated human polynucleotides encoding polypeptides, useful for the
 XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
 XX

XX Claim 20; Page 260; 1217pp; English.
 PS
 XX
 CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnerary;
 CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.
 XX
 XX Sequence 260 AA;
 SQ

Query Match 100.0%; Score 49; DB 22; Length 260;
 Best Local Similarity 100.0%; Pred. No. 0.088;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EPIDGKGL 9
 IIIIIII
 DB 126 epiddgkgl 134

RESULT 7
 ABB66320
 ID ABB66320 standard; Protein: 299 AA.
 XX
 XX ABB66320;
 XX
 XX 26-MAR-2002 (first entry)
 XX
 XX Drosophila melanogaster polypeptide SEQ ID NO 25752.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 XX
 XX 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 XX
 XX 23-MAR-2000; 2000US-191637P.
 PR
 XX 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI; 2001-656860/75.
 DR
 XX N-PSDB; ABL10423.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Disclosure; SEQ ID NO 25752; 21pp + Sequence Listing; English.
 PS

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 299 AA;

Query Match 77.6%; Score 38; DB 22; Length 299;
 Best Local Similarity 75.0%; Pred. No. 17;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EPIDGKG 8
 ||:||||
 Db 133 epvddgpg 140

RESULT 8

ABB67420
 ID ABB67420 standard; Protein; 299 AA.

XX AC ABB67420;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 29052.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL11523.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX PS Disclosure; SEQ ID NO 29052; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 299 AA;

Query Match 77.6%; Score 38; DB 22; Length 299;
 Best Local Similarity 75.0%; Pred. No. 17;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EPIDGKG 8
 ||:||||
 Db 133 epvddgpg 140

RESULT 9

ABB60369
 ID ABB60369 standard; Protein; 2167 AA.

XX AC ABB60369;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 7899.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL04472.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX PS Disclosure; SEQ ID NO 7899; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 2167 AA;

Query Match 77.6%; Score 38; DB 22; Length 2167;
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIDGKG 7

Db 926 epvddgk 932

RESULT 10
ABG24503
ID ABG24503 standard; Protein: 98 AA.

XX AC ABG24503;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #24494.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.
XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB; AAS88690.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.

XX PS Claim 20; SEQ ID NO 54862; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 98 AA;

Query Match 71.4%; Score 35; DB 22; Length 98;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EPIDGKGL 9
| | | | | | | |
Db 66 eviddgrgm 74

RESULT 12

RESULT 11

ABG29553
ID ABG29553 standard; Protein: 127 AA.

XX AC ABG29553;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #29544.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB; AAS93740.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.

XX PS Claim 20; SEQ ID NO 59912; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 127 AA;

Query Match 71.4%; Score 35; DB 22; Length 127;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EPIDGKGL 9
| | | | | | | |
Db 67 eviddgrgm 75

AAU04903
ID AAU04903 standard; Protein: 296 AA.
XX AAU04903;
XX
XX
XX 26-SEP-2001 (first entry)
XX
XX
XX Micromonospora everninomicin biosynthetic enzyme ORF3.
XX
XX Everninomicin; antibiotic; bottle-neck gene; orthomycin;
KW fermentation; ORF3.
XX
XX Micromonospora carbonacea var. africana.
OS
XX
XX WO200151639-A2.
PN
XX
XX 19-JUL-2001.
PD
XX
XX 12-JAN-2001; 2001WO-US01187.
PF
XX
XX 12-JAN-2000; 2000US-0175751.
PR
XX
XX (SCHE) SCHERING CORP.
PA
XX
XX Hosted TJ, Horan AC, Wang TX;
PI
XX
XX WPI; 2001-442147/47.
DR
XX
XX N-PSDB; AAS08699.
XX
XX New nucleic acid molecules encoding everninomicin pathway gene
PT products, useful for improving yields of everninomicin, to produce new
PT everninomicin and as probes to identify homologous sequences -
XX
XX Claim 19; Fig 12; 109pp; English.
XX
XX The sequence comprises one of 98 enzymes of the everninomicin
CC antibiotic biosynthetic pathway, ORF3. A vector comprising a
CC M. carbonacea everninomicin biosynthetic pathway resistance gene product
CC is useful for selecting for a transfected or transformed host cell. An
CC integrative version of the vector is useful for introducing a
CC everninomicin pathway gene (a bottle-neck gene) into an actinomycete of
CC the genus Micromonospora. The DNA encoding the biosynthetic proteins is
CC useful for synthesizing novel everninomicin-related compounds, arising
CC from modifications of the DNA sequence designed to change glycosyl and
CC modified orsellinic acid groups contained in everninomicin, for
CC expressing functional or mutant everninomicin biosynthetic enzyme for
CC evaluation, diagnosis and preferably everninomicin biosynthesis of
CC other secondary metabolic products, improving the yield of everninomicin or
CC and to produce novel everninomicins and also as a hybridisation probe to
CC identify homologous sequences. The encoded polypeptides are useful for
CC combinatorial biosyntheses to generate libraries of orthomycins, e.g.
CC everninomicin analogues/homologues and drug discovery. The
CC DNA encoding the integrase allows for increasing a given gene dosage. The
CC integrative vector can be used to permanently integrate copies of a
CC heterologous gene of choice into chromosomes of different hosts and to
CC integrate genes which increase the yield of known products or to generate
CC novel products such as hybrid antibiotics or other novel secondary
CC metabolites. The vector can also be used to integrate antibiotic
CC resistance genes in order to carry out bioconversions with compounds to
CC which the strain is normally sensitive and is thus useful in fermentation
CC processes involving e.g. Streptomyces antibiotics.
XX
XX Sequence 296 AA;
SQ

AAU57477
ID AAU57477 standard; Protein: 414 AA.
XX AAU57477;
XX
XX 13-FEB-2002 (first entry)
XX
XX
XX Propionibacterium acnes immunogenic protein #18373.
DE
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
XX
XX Propionibacterium acnes.
OS
XX
XX WO200181581-A2.
PN
XX
XX 01-NOV-2001.
PD
XX
XX 20-APR-2001; 2001WO-US12865.
PF
XX
XX 21-APR-2000; 2000US-199047P.
PR
XX
XX 02-JUN-2000; 2000US-208841P.
PR
XX
XX 07-JUL-2000; 2000US-216747P.
PR
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI; 2001-616774/71.
DR
XX
XX N-PSDB; AAS59583.
DR
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
XX Example 1; SEQ ID No 18672; 1059pp; English.
PS
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 414 AA;
SQ

Query Match 71.4%; Score 35; DB 22; Length 296;
Best Local Similarity 71.4%; Pred. No. 68;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 2 PIDGKG 8
Db 189 plddgrg 195
:|||||

Query Match 71.4%; Score 35; DB 22; Length 414;
Best Local Similarity 71.4%; Pred. No. 1e+02; 0; Indels 0; Gaps 0;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EPIDGKG 7
Db 223 dpvddgk 229
:|||||

RESULT 14

ABG29555
ID ABG29555 standard; Protein: 481 AA.

XX AC ABG29555;
XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #29546.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS93742.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity

XX PS Claim 20: SEQ ID No 59914; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 481 AA;

Query Match 71.4%; Score 35; DB 22; Length 481;

Best Local Similarity 66.7%; Pred. No. 1.2e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EPIDCKGL 9

1 111111

DB 417 evidgkgl 425

RESULT 15

AAU41020
ID AAU41020 standard; Protein: 511 AA.

XX AC AAU41020;

XX DT 13-FEB-2002 (first entry)

XX DE Propionibacterium acnes immunogenic protein #1916.

XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX KW dermatological; osteopathic; neuroprotectant.

XX OS Propionibacterium acnes.

XX PN WO200181581-A2.

XX PD 01-NOV-2001.

XX PF 20-APR-2001; 2001WO-US12865.

XX PR 21-APR-2000; 2000US-195047P.

XX PR 02-JUN-2000; 2000US-208841P.

XX PR 07-JUL-2000; 2000US-216747P.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX DR WPI; 2001-616774/71.

XX DR N-PSDB; AAS59513.

XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris

XX PS Example 1; SEQ ID No 2215; 1069pp; English.

XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX CC polypeptides. The proteins and their associated DNA sequences are used in
XX CC the treatment, prevention and diagnosis of medical conditions caused by
XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX CC P. acnes is also involved in infections of bone, joints and the central
XX CC nervous system, however it is particularly involved in the inflammatory
XX CC lesions associated with acne vulgaris. A method for detecting the
XX CC presence or absence of P. acnes in a patient comprises contacting a
XX CC sample with a binding agent that binds to the proteins of the invention
XX CC and determining the amount of bound protein in the sample. The
XX CC polypeptides may be used as antigens in the production of antibodies
XX CC specific for P. acnes proteins. These antibodies can be used to
XX CC downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA).

XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 511 AA;

Query Match 71.4%; Score 35; DB 22; Length 511;

Best Local Similarity 75.0%; Pred. No. 1.3e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PIDCKGL 9

1 111111

DB 29 plddgkdl 36

Wed Aug 7 05:46:35 2002

us-10-020-139-2_copy_115_123.rag

Page 9

Search completed: August 6, 2002, 16:52:54
Job time: 336 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 16:53:55 ; Search time 51.22 Seconds
(without alignments)
4.292 Million cell updates/sec

Title: US-10-020-139-2_COPY_115_123

Perfect score: 49

Sequence: 1 EPIDCKGL 9

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	67.3	396	US-08-501-126-2	Sequence 2, Appl
2	33	67.3	397	US-08-282-197C-55	Sequence 55, Appl
3	33	67.3	603	US-08-687-865A-2	Sequence 2, Appl
4	33	67.3	603	US-09-043-711-2	Sequence 2, Appl
5	32	65.3	58	US-08-470-179-15	Sequence 15, Appl
6	32	65.3	267	US-08-718-905-3	Sequence 3, Appl
7	32	65.3	267	US-09-550-497-3	Sequence 3, Appl
8	32	65.3	267	US-09-147-992-3	Sequence 3, Appl
9	32	65.3	271	US-09-323-427-9	Sequence 9, Appl
10	32	65.3	419	US-09-011-197-4	Sequence 4, Appl
11	32	65.3	658	US-08-409-995-5	Sequence 5, Appl
12	32	65.3	658	US-08-685-467-5	Sequence 5, Appl
13	32	65.3	658	US-08-913-942-5	Sequence 5, Appl
14	32	65.3	1098	US-08-409-995-2	Sequence 2, Appl
15	32	65.3	1098	US-08-685-467-2	Sequence 2, Appl
16	32	65.3	1098	US-09-377-155-32	Sequence 32, Appl
17	32	65.3	1098	US-08-913-942-2	Sequence 32, Appl
18	32	65.3	1098	US-09-669-974-32	Sequence 32, Appl
19	32	65.3	1098	US-09-268-347-44	Sequence 44, Appl
20	32	65.3	1525	US-09-396-651B-1	Sequence 1, Appl
21	32	65.3	1751	US-09-136-574A-44	Sequence 44, Appl
22	31	63.3	158	US-08-470-179-14	Sequence 14, Appl
23	31	63.3	159	US-08-858-207A-509	Sequence 509, App
24	31	63.3	304	US-09-232-200-100	Sequence 100, App
25	31	63.3	304	US-09-232-197-100	Sequence 100, App
26	31	63.3	304	US-09-232-201-100	Sequence 100, App
27	31	63.3	346	US-08-476-254-2	Sequence 2, Appl

Query Match

67.3%; Score 33; DB 4; Length 396;

Sequence 10, Appl
Patent No. 547933
Patent No. 547933
Sequence 2, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 26, Appl
Sequence 26, Appl
Sequence 26, Appl
Sequence 26, Appl
Sequence 26, Appl
Sequence 8, Appl
Sequence 8, Appl
Sequence 8, Appl
Sequence 8, Appl
Sequence 331, App
Sequence 3, Appl
Sequence 2, Appl
Sequence 28, Appl

28 31 63.3 346 2 US-08-476-254-10
29 31 63.3 346 6 547933-2
30 31 63.3 346 6 547933-7
31 31 63.3 363 3 US-09-046-086-2
32 31 63.3 438 1 US-08-480-604A-23
33 31 63.3 438 2 US-08-405-496A-23
34 31 63.3 438 4 US-08-915-136-23
35 31 63.3 462 1 US-08-480-604A-26
36 31 63.3 462 2 US-08-405-496A-26
37 31 63.3 462 4 US-08-915-136-26
38 31 63.3 503 1 US-08-245-294-8
39 31 63.3 503 1 US-08-474-499-8
40 31 63.3 503 1 US-08-307-279A-8
41 31 63.3 503 5 PCT-US95-06211-8
42 31 63.3 596 4 US-09-199-637A-331
43 31 63.3 645 1 US-07-779-172A-3
44 31 63.3 883 1 US-08-106-433A-2
45 31 63.3 1296 1 US-08-480-604A-28

ALIGNMENTS

RESULT 1
US-08-501-126-2
; Sequence 2, Application US/08501126
; Patent No. 6140095
; GENERAL INFORMATION:
; APPLICANT: Van Solingen, Pieter
; APPLICANT: Williams, Diane P.
; APPLICANT: Iverson, Sara
; APPLICANT: Farrell, Roberta L.
; APPLICANT: Herbes, Wilhelmina T.
; APPLICANT: Van Der Kleij, Wilhelmus A.
; APPLICANT: Herweijer, Margaretha A.
; APPLICANT: Van Beckhoven W.C., Rudolf F.
; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Jones, Brian E.
; TITLE OF INVENTION: ALCALI-TOLERANT XYLANASES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,126
; FILING DATE: 29-DEC-1995
; CLASSIFICATION: 425
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 4615-0057.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-501-126-2

Best Local Similarity 77.8%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EPIDGKGL 9
| | | | | | | |
Db 195 EVIDGGGL 203

RESULT 2
US-08-282-197C-55
; Sequence 55, Application US/08282197C
; Patent No. 5871730
; GENERAL INFORMATION:
; APPLICANT: Brzezinski, Kyszard
; APPLICANT: Dery, Claude V
; APPLICANT: Beaulieu, Carole
; TITLE OF INVENTION: Thermostable Xylanase DNA, Protein and
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,197C
; FILING DATE: 29-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbal, Michele A
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 1050.0410000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2540
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; TOPOLOGY: both
US-08-282-197C-55

Query Match 67.3%; Score 33; DB 2; Length 397;
Best Local Similarity 77.8%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EPIDGKGL 9
| | | | | | | |
Db 195 EVIDGGGL 203

RESULT 3
US-08-687-865A-2
; Sequence 2, Application US/08687865A
; Patent No. 5955596
; GENERAL INFORMATION:
; APPLICANT: Jones, Kevin F.
; APPLICANT: Zagursky, Robert J.
; TITLE OF INVENTION: The Nuca Protein of Haemophilus
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza

CITY: Wayne
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07470
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,865A
FILING DATE: 26-JUL-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Gordon, Alan M.
REGISTRATION NUMBER: 30,637
REFERENCE/DOCKET NUMBER: 33,250-00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-831-3244
TELEFAX: 201-831-3305
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 603 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-687-865A-2

Query Match 67.3%; Score 33; DB 2; Length 603;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EPIDGK 7
| | | | |
Db 525 EPIDDK 531

RESULT 4
US-09-043-711-2
; Sequence 2, Application US/09043711
; Patent No. 6221365
; GENERAL INFORMATION:
; APPLICANT: Jones, Kevin F.
; APPLICANT: Zagursky, Robert J.
; TITLE OF INVENTION: The Nuca Protein of Haemophilus
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07470
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/043,711
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/687,865
FILING DATE: 26-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gordon, Alan M.
REGISTRATION NUMBER: 30,637
REFERENCE/DOCKET NUMBER: 33,250-00
TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-831-3244
 TELEFAX: 201-831-3305
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 603 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-043-711-2

Query Match 67.3%; Score 33; DB 4; Length 603;
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EPIDGK 7
 |||||
 DB 525 EPIDNK 531

RESULT 5
 US-08-470-179-15
 Sequence 15, Application US/08470179
 Patent No. 5645994

GENERAL INFORMATION:
 APPLICANT: Huang Ph.D. Wai Mun
 TITLE OF INVENTION: Method and Compositions for
 IDENTIFICATION OF SPECIES IN A SAMPLE
 NUMBER OF SEQUENCES: 207
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Trask, Britt and Rossa
 STREET: P.O. Box 2550
 CITY: Salt Lake City
 STATE: Utah
 COUNTRY: USA
 ZIP: 84110

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/470,179
 FILING DATE:

CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Sweigert Ph.D. Susan E.
 REGISTRATION NUMBER: 36,289
 REFERENCE/DOCKET NUMBER: 2601
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 801-532-1922
 TELEFAX: 801-531-9168

INFORMATION FOR SEQ ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 58 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Borrelia burgdorferi
 US-08-470-179-15

Query Match 65.3%; Score 32; DB 1; Length 58;
 Best Local Similarity 71.4%; Pred. No. 19;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 IDGKGL 9
 |||||
 DB 51 IDGRI 57

RESULT 6

US-08-718-905-3
 Sequence 3, Application US/08718905
 Patent No. 6063756
 GENERAL INFORMATION:
 APPLICANT: Donovan, William P.
 APPLICANT: Slaney, Judith C.
 TITLE OF INVENTION: BACILLUS THURINGIENSIS CRYET33 AND CRYET34
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P.O. Box 4433
 CITY: Houston
 STATE: Texas
 COUNTRY: United States of America
 ZIP: 77210

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/718,905
 FILING DATE: Concurrently Herewith
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Kitchell, Barbara S.
 REGISTRATION NUMBER: 33,928
 REFERENCE/DOCKET NUMBER: MOBT:003
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3000
 TELEFAX: (512) 474-7577

INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 267 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-718-905-3

Query Match 65.3%; Score 32; DB 3; Length 267;
 Best Local Similarity 71.4%; Pred. No. 95;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PIDGKG 8
 |||||
 DB 235 PLDDNKG 241

RESULT 7

US-09-550-497-3
 Sequence 3, Application US/09550497
 Patent No. 6248536
 GENERAL INFORMATION:
 APPLICANT: Donovan, William P.
 APPLICANT: Slaney, Judith C.
 TITLE OF INVENTION: BACILLUS THURINGIENSIS CRYET33 AND CRYET34
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P.O. Box 4433
 CITY: Houston
 STATE: Texas
 COUNTRY: United States of America
 ZIP: 77210

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/09/550,497
FILING DATE: 14-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/718,905
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: MORT:003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 267 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-550-497-3

Query Match 65.3%; Score 32; DB 4; Length 267;
Best Local Similarity 71.4%; Pred. No. 95;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PIDGKG 8
I:|I|I|I|
DB 235 PLDDNKG 241

RESULT 8
US-09-147-992-3
Sequence 3, Application US/09147992
Patent No. 6326351
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: BACILLUS THURINGIENSIS CRYET33 AND CRYET34
TITLE OF INVENTION: COMPOSITIONS AND USES THEREOF
NUMBER OF SEQUENCES: 10
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/147,992
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/718,905
FILING DATE: 24-SEP-1996
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 267 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-147-992-3

Query Match 65.3%; Score 32; DB 4; Length 267;
Best Local Similarity 71.4%; Pred. No. 95;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PIDGKG 8
I:|I|I|I|
DB 235 PLDDNKG 241

RESULT 9
US-09-323-427-9
Sequence 9, Application US/09323427
Patent No. 6248329
GENERAL INFORMATION:
APPLICANT: Chandrashekar, Ramaswamy
TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
TITLE OF INVENTION: Molecules, and Uses Thereof
FILE REFERENCE: HW-8
CURRENT APPLICATION NUMBER: US/09/323,427
CURRENT FILING DATE: 1999-06-01
EARLIER APPLICATION NUMBER: 60/087,435
EARLIER FILING DATE: 1998-06-01
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 9
LENGTH: 271
TYPE: PRT
ORGANISM: *Dirofilaria immitis*
US-09-323-427-9

Query Match 65.3%; Score 32; DB 4; Length 271;
Best Local Similarity 83.3%; Pred. No. 97;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 IDGKG 8
:|I|I|I|I|
DB 80 VDDGKG 85

RESULT 10
US-09-011-197-4
Sequence 4, Application US/09011197
Patent No. 6171789
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: No. 6171789el Insertion Sequence from a Virulent
TITLE OF INVENTION: Isolate of Burkholderia Cepacia, and Diagnostic and
TITLE OF INVENTION: Identification Procedures Based Thereon.
NUMBER OF SEQUENCES: 16
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/011,197
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/002,398
FILING DATE: 17-AUG-1995
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 419 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: *Burkholderia cepacia*
INDIVIDUAL ISOLATE: ET12/cblA
US-09-011-197-4

Query Match 65.3%; Score 32; DB 4; Length 419;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

QY      2 PIDDGKGL 9
      1::|1111
Db      10 PVEPGKGL 17

RESULT 11
US-08-409-995-5
; Sequence 5, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; TITLE OF INVENTION: St. Gene III, Joseph W.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 658 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-409-995-5

Query Match 65.3%; Score 32; DB 1; Length 658;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 EPIDGKGL 9
      1::|1111
Db      350 EDADEGKGL 358

RESULT 12
US-08-685-467-5
; Sequence 5, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 658 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-685-467-5

Query Match 65.3%; Score 32; DB 3; Length 658;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 EPIDGKGL 9
      1::|1111
Db      350 EDADEGKGL 358

RESULT 13
US-08-913-942-5
; Sequence 5, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA: PCT/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 658 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-913-942-5

Query Match 65.3%; Score 32; DB 4; Length 658;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EPIDGKGL 9
| 1:1111
DB 350 EDADEGKGL 358

RESULT 14
US-08-409-995-2
Sequence 2, Application US/08409995
Patent No. 5646259
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen I.
APPLICANT: St. Gene III, Joseph W.
TITLE OF INVENTION: Haemophilus Adhesion Proteins
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,995
FILING DATE: 24-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1098 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
US-08-409-995-2

Query Match 65.3%; Score 32; DB 1; Length 1098;
Best Local Similarity 66.7%; Pred. No. 4.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EPIDGKGL 9
| 1:1111
DB 350 EDADEGKGL 358

RESULT 15
US-08-685-467-2
Sequence 2, Application US/08685467
Patent No. 6060059
GENERAL INFORMATION:
APPLICANT: St. Gene III, Joseph W.
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,467
FILING DATE: 22-JUL-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1098 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-685-467-2

Query Match 65.3%; Score 32; DB 3; Length 1098;
Best Local Similarity 66.7%; Pred. No. 4.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EPIDGKGL 9
| 1:1111
DB 350 EDADEGKGL 358

Search completed: August 6, 2002, 16:53:56
Job time: 288 sec

Wed Aug , 7 05:46:36 2002

us-10-020-139-2_copy_115_123.rai

Page 7

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 06:43:38 ; Search time 14.84 seconds
(without alignments)
51,800 Million cell updates/sec

Title: US-10-020-139-2_copy_159_166

Perfect score: 46

Sequence: 1 ETDPQTHQ 8

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	38	82.6	274	2	5,10-methylenetetrahydrofolate reductase (EC 1.7.99.5) [imported] - Brucella meliter
2	37	80.4	445	2	GDP dissociation inhibitor protein - rice
3	37	80.4	445	2	GDP dissociation inhibitor protein - rice
4	36	78.3	880	2	hypothetical protein
5	35	76.1	250	2	probable transcript
6	35	76.1	425	2	aminotransferase X
7	35	76.1	444	2	GDP dissociation inhibitor protein
8	35	76.1	532	2	methionine--tRNA ligase
9	35	76.1	1240	2	DNA polymerase III
10	35	76.1	1240	2	DNA polymerase III
11	34	73.9	443	2	GDP dissociation inhibitor protein
12	34	73.9	670	1	arachidonate 5-lipoxygenase
13	34	73.9	674	2	arachidonate 5-lipoxygenase
14	34	73.9	1223	2	hypothetical protein
15	33	71.7	92	1	Sugar fermentation
16	33	71.7	92	2	Sugar fermentation
17	33	71.7	122	2	conserved hypothetical protein
18	33	71.7	321	2	hypothetical protein
19	33	71.7	380	2	probable omega-3 fatty acid desaturase
20	33	71.7	422	2	serine hydroxymethyltransferase
21	33	71.7	422	2	serine hydroxymethyltransferase
22	33	71.7	428	2	conserved hypothetical protein
23	33	71.7	428	2	hypothetical protein
24	33	71.7	436	2	hypothetical protein
25	33	71.7	470	2	UDP-N-acetylmuramoyl transferase precursor
26	33	71.7	495	2	hypothetical protein
27	33	71.7	706	2	hypothetical protein
28	33	71.7	990	2	hypothetical protein
29	33	71.7	1329	2	cobalamin biosynthesis

30 32 59.6 109 2 D97991 hypothetical prote
31 32 59.6 146 2 F75400 hypothetical prote
32 32 59.6 207 2 T31239 traaE protein homol
33 32 59.6 233 2 T08326 hypothetical prote
34 32 59.6 236 2 D85096 probable DNA-bind
35 32 59.6 239 2 T02984 myb-related protei
36 32 59.6 249 1 S68688 myb-related protei
37 32 59.6 274 1 JQ0957 myb-related protei
38 32 59.6 294 2 T09879 myb-related protei
39 32 59.6 416 2 T19149 hypothetical prote
40 32 59.6 466 2 E70112 asparagine--tRNA l
41 32 59.6 522 2 S75491 hypothetical prote
42 32 59.6 567 2 AC0143 choline dehydrogen
43 32 59.6 603 2 F70651 probable atsf prot
44 32 59.6 699 2 C43674 US4 protein - huma
45 32 59.6 739 2 S15727 cellulase (EC 3.2.

ALIGNMENTS

RESULT 1
A13321
5,10-methylenetetrahydrofolate reductase (EC 1.7.99.5) [imported] - Brucella meliter
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #Sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: A13321
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivan
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; L
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A>Title: The genome sequence of the facultative intracellular pathogen Brucella meli
A:Reference number: A03252; PMID:11756688
A:Accession: A13321
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-274 <KUR>
A:Cross-references: GB:AE008917; PID:AAU51740.1; PID:g17982477; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME10559
A:Map position: 1
C:Superfamily: 5,10-methylenetetrahydrofolate reductase (FADH2)
C:Keywords: oxidoreductase

Query Match 82.6%; Score 38; DB 2; Length 274;
Best Local Similarity 75.0%; Pred. No. 6.9;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDPQTHQ 8
: |||||
Db 214 DNDPQTHQ 221

RESULT 2
T02030

GDP dissociation inhibitor protein - rice
C:Species: Oryza sativa (rice)
C>Date: 26-Feb-1999 #Sequence_revision 26-Feb-1999 #text_change 26-May-2000
R:Kim, C.Y.; Cheon, S.Y.; Jeong, S.Y.; Cheong, Y.H.; Choe, M.S.; Park, H.C.; Cho, M
submitted to the EMBL Data Library, August 1997
A:Reference number: Z14502
A:Accession: T02030
A>Status: preliminary; translated from GR/EMBL/DDB
A:Molecule type: mRNA
A:Residues: 1-445 <KIM>
A:Cross-references: EMBL:AF016896; NID:g2384757; PID:g2384758
C:Genetics:
A:Gene: osGDII
C:Superfamily: human GDP dissociation inhibitor XAP-4

Query Match 80.4%; Score 37; DB 2; Length 445;
 Best Local Similarity 75.0%; Pred. No. 18;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDPOTHQ 8
 | | | | |
 Db 151 EADPKTHQ 158

RESULT 3

T02032
 GDP dissociation inhibitor protein - rice
 C:Species: Oryza sativa (rice)
 C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 26-May-2000
 C:Accession: T02032
 R:Kim, C.Y.; Cheon, S.Y.; Jeong, S.Y.; Cheong, Y.H.; Choe, M.S.; Park, H.C.; Cho, M.J.
 submitted to the EMBL Data Library, August 1997
 A:Reference number: Z14502
 A:Accession: T02032
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-445 <REV>
 A:Cross-references: EMBL:AF016897; NID:g2384759; PID:g2384750
 C:Genetics:
 A:Gene: GDI2
 C:Superfamily: human GDP dissociation inhibitor XAP-4

Query Match 80.4%; Score 37; DB 2; Length 445;
 Best Local Similarity 75.0%; Pred. No. 18;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDPOTHQ 8
 | | | | |
 Db 151 EADPKTHQ 158

RESULT 4

T04523
 hypothetical protein F16A16.130 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
 C:Accession: T04523
 R:Bevan, M.; Brandt, P.; Dose, S.; Jarke, D.; Scharfe, M.; Schöni, O.; Hoheisel, J.; Mew
 submitted to the Protein Sequence Database, February 1999
 A:Reference number: Z15376
 A:Accession: T04523
 A:Molecule type: DNA
 A:Residues: 1-880 <REV>
 A:Cross-references: EMBL:AL035353
 A:Experimental source: Cultivar Columbia; BAC clone F16A16
 C:Genetics:
 A:Map position: 4
 A:Introns: 660/1
 A:Note: F16A16.130

Query Match 78.3%; Score 36; DB 2; Length 880;
 Best Local Similarity 87.5%; Pred. No. 59;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDPOTHQ 8
 | | | | |
 Db 46 ETLPOTHQ 53

RESULT 5

T48145
 probable transcription factor - Arabidopsis thaliana
 N:Alternate names: protein T4C9.190
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: T48145

R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.;
 submitted to the Protein Sequence Database, June 1999
 A:Reference number: Z24485
 A:Accession: T48145
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-250 <REV>
 A:Cross-references: EMBL:AL080318
 A:Experimental source: cultivar Columbia; BAC clone T4C9
 C:Genetics:
 A:Map position: 4
 A:Introns: 2/2; 52/2
 A:Note: T4C9.190

Query Match 76.1%; Score 35; DB 2; Length 250;
 Best Local Similarity 75.0%; Pred. No. 24;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ETDPOTHQ 8
 | | | | |
 Db 84 EIDPSTHQ 91

RESULT 6

C82564
 aminotransferase XF2396 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: C82564
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Se
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: C82564
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-425 <SIM>
 A:Cross-references: GB:AE004048; G9:AE003849; NID:g9107566; PIDN:AAF85195.1; GSPDB:G
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
 as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraija, J.S.; Franca, S.C.; Franco, M.C.; F
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; L
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martin
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Saw
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Sil
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF2396
 C:Superfamily: aspartate transaminase

Query Match 76.1%; Score 35; DB 2; Length 425;
 Best Local Similarity 85.7%; Pred. No. 42;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TDPOTHQ 8
 | | | | |
 Db 71 TDPYTHQ 77

RESULT 7

T01782
 GDP dissociation inhibitor - common tobacco
 C:Species: Nicotiana tabacum (common tobacco)

C>Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 26-May-2000
 C:Accession: T01782
 R: Ezaki, B.; Koyanagi, M.; Gardner, R.C.; Matsumoto, H.
 submitted to the EMBL Data Library, July 1997
 A:Reference number: Z14424
 A:Accession: T01782
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-444 <EZA>
 A:Cross-references: EMBL:AF012823; NID:g2501849; PID:g2501850
 A:Experimental source: Cullivar Samsun
 C:Genetics:
 A:Gene: GDI
 C:Function:
 A:Description: inhibits dissociation of GDP from GTP binding proteins
 C:Superfamily: human GDP dissociation inhibitor XAP-4

Query Match 76.1%; Score 35; DB 2; Length 444;
 Best Local Similarity 62.5%; Pred. No. 44;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDPTQTHQ 8
 I:|||||
 Db 151 ESDPKTHE 158

RESULT 8
 S76525
 methionine--tRNA ligase (EC 6.1.1.10) - Synecchocystis sp. (strain PCC 6803)
 C:Species: Synecchocystis sp.
 A:Variety: PCC 6803
 C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
 C:Accession: S76525
 R: Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3: 109-136 1996
 A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis
 S.
 A:Reference number: S74322; MUID:97061201
 A:Accession: S76525
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-532 <RAN>
 A:Cross-references: EMBL:D64002; GB:AB001339; NID:g1001612; PIDN:BAA10371.1; PID:g100164
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Superfamily: methionine--tRNA ligase
 C:Keywords: ligase

Query Match 76.1%; Score 35; DB 2; Length 532;
 Best Local Similarity 85.7%; Pred. No. 53;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDPTQTH 7
 I:|||||
 Db 68 ELDPQTH 74

RESULT 9
 G86573
 DNA polymerase III alpha [imported] - Chlamydomophila pneumoniae (strain J138)
 C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C:Accession: G86573
 R: Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
 Nucleic Acids Res. 28: 2311-2314 2000
 A>Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
 A:Reference number: A86491; MUID:20330349
 A:Accession: G86573
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1240 <STO>

A:Cross-references: GB:BA000008; NID:g8979038; PIDN:BAA98873.1; GSPDB:GN00142
 A:Experimental source: strain J138
 C:Genetics:
 A:Gene: dnaE
 C:Superfamily: DNA-directed DNA polymerase III alpha chain

Query Match 76.1%; Score 35; DB 2; Length 1240;
 Best Local Similarity 75.0%; Pred. No. 1.3e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ETDPTQTHQ 8
 I:|||||
 Db 525 ETDPLHQ 532

RESULT 10
 B72050
 DNA polymerase III, alpha chain [imported] - Chlamydomophila pneumoniae (strain
 C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: B72050; E81616
 R: Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Ollinger, L.; Grimwood
 Nature Genet. 21: 385-389, 1999
 A>Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606
 A:Accession: B72050
 A:Molecule type: DNA
 A:Residues: 1-1240 <ARN>
 A:Cross-references: GB:AF001649; GB:AE001363; NID:g4376963; PIDN:AAI8805.1; PID:g4
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hic
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salz
 Nucleic Acids Res. 28: 1397-1406 2000
 A>Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR
 A:Reference number: A81500; MUID:20150255
 A:Accession: E81616
 A:Molecule type: DNA
 A:Residues: 1-1240 <REA>
 A:Cross-references: GB:AE002171; GB:AE002161; NID:g7189009; PIDN:AAF37967.1; PID:g7
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: dnaE; CP0081
 C:Superfamily: DNA-directed DNA polymerase III alpha chain

Query Match 76.1%; Score 35; DB 2; Length 1240;
 Best Local Similarity 75.0%; Pred. No. 1.3e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ETDPTQTHQ 8
 I:|||||
 Db 525 ETDPLHQ 532

RESULT 11
 T10801
 GDP dissociation inhibitor GDI1 - Volvox carteri f. nagariensis
 C:Species: Volvox carteri f. nagariensis
 C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 26-May-2000
 C:Accession: T10801
 R: Beyer, K.; Fabry, S.
 submitted to the EMBL Data Library, July 1996
 A:Description: Identification and characterization of a lower plant Ypt/Rab guanosyl
 A:Reference number: Z17156
 A:Accession: T10801
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-443 <BEY>
 A:Cross-references: EMBL:D62866; NID:g1572517; PID:g1572518
 A:Experimental source: strain HK10
 C:Genetics:
 A:Gene: GDI1

C:Function:
A:Description: Inhibits dissociation of GDP from GTP binding proteins
C:Superfamily: human GDP dissociation inhibitor XAP-4

Query Match 73.9%; Score 34; DB 2; Length 443;
Best Local Similarity 62.5%; Pred. No. 68;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ETDPTQHQ 8
|:|:|:|:
Db 151 ESDPKTHK 158

RESULT 12

A30882

arachidonate 5-lipoxygenase (EC 1.13.11.34) - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A30882

R:Balcarek, J.M.; Theisen, T.W.; Cook, M.N.; Varrichio, A.; Hwang, S.M.; Strohsacker, M.

J. Biol. Chem. 263, 13937-13941, 1998

A:Title: Isolation and characterization of a cDNA clone encoding rat 5-lipoxygenase.

A:Reference number: A30882; MUID:88330933

A:Accession: A30882

A:Molecule type: mRNA

A:Residues: 1-670 <BAL>

A:Cross-references: GB:J03960; NID:q205228; PIDN:AAA41538.1; PID:g205229

C:Superfamily: arachidonate 5-lipoxygenase

C:Keywords: oxidoreductase

Query Match 73.9%; Score 34; DB 1; Length 670;
Best Local Similarity 75.0%; Pred. No. 1;le=02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ETDPTQHQ 8
|:|:|:|:
Db 296 KTDPTCHQ 303

RESULT 13

I49479

arachidonate 5-lipoxygenase - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 11-Jun-1999

C:Accession: I49479

R:Chen, X.S.; Naumann, T.A.; Kurre, U.; Jenkins, N.A.; Copeland, N.G.; Funk, C.D.

J. Biol. Chem. 270, 17993-17999, 1995

A:Title: cDNA cloning, expression, mutagenesis, intracellular localization, and gene chr

A:Reference number: A57186; MUID:95355399

A:Accession: I49479

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-674 <RES>

A:Cross-references: GB:L42198; NID:g886332; PIDN:AAC37673.1; PID:g886333

C:Genetics:

A:Gene: Alox5

C:Superfamily: arachidonate 5-lipoxygenase

Query Match 73.9%; Score 34; DB 2; Length 674;
Best Local Similarity 75.0%; Pred. No. 1;le=02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ETDPTQHQ 8
|:|:|:|:
Db 297 KTDPTCHQ 304

RESULT 14

T17345

hypothetical protein DKFZp586M1824.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T17345

R:Duesterhoef, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

Submitted to the Protein Sequence Database, September 1999

A:Reference number: Z18727

A:Accession: T17345

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1223 <DUE>

A:Cross-references: EMBL:AL117665

A:Experimental source: adult uterus; clone DKFZp586M1824

C:Genetics:

A:Note: DKFZp586M1824.1

Query Match 73.9%; Score 34; DB 2; Length 1223;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ETDPTQHQ 8
|:|:|:|:
Db 1040 ETDPTSQ 1047

RESULT 15

BVECNP

Sugar fermentation stimulation protein B (NER-like protein) - Escherichia coli

N:Alternate names: ner-like protein

C:Species: Escherichia coli

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 27-Nov-2001

C:Accession: JY0027; F65109

R:Choi, Y.L.; Nishida, T.; Kawamukai, M.; Utsumi, R.; Sakai, H.; Komano, T.

J. Bacteriol. 171, 5223-5225, 1989

A:Title: Cloning and sequencing of an Escherichia coli gene, nlp, highly homologous

A:Reference number: JY0027; MUID:89359178

A:Accession: JY0027

A:Molecule type: DNA

A:Residues: 1-92 <CHO>

A:Cross-references: GB:X68873; NID:g42128; PIDN:CAA48736.1; PID:g42129

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: F65109

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-92 <BLAT>

A:Cross-references: GB:AF000399; GB:U00096; NID:g2367201; PIDN:AAC76220.1; PID:g178

A:Experimental source: strain K-12, substrain MG1655

C:Comment: This protein is involved in positive regulation of the metabolism of suga

C:Genetics:

A:Gene: nlp

A:Map position: 69 min

C:Superfamily: phage D108 DNA-binding protein

C:Keywords: DNA binding; transcription regulation

F:50-68/Region: DNA binding #status predicted

Query Match 71.7%; Score 33; DB 1; Length 92;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DPTQHQ 8
|:|:|:|:
Db 71 DPTQHE 76

Search completed: August 7, 2002, 06:44:03

Job time: 25 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: August 7, 2002, 06:44:08 ; Search time 10.33 Seconds
(without alignments)
29,986 Million cell updates/sec

Title: US-10-020-139-2_COPY_159_166

Perfect score: 46
Sequence: 1 FTDPTQTHQ 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	76.1	532	1 SYM_SYNY3	Q55729 synchocyst
2	35	76.1	1240	1 DP3A_CHLPN	Q927n8 chlamydia p
3	34	73.9	1778	1 FXV5_HUMAN	Q96db9 homo sapien
4	34	73.9	672	1 LOXS_MESAU	P51399 mesocricetu
5	34	73.9	672	1 LOXS_RAT	P12527 rattus norv
6	34	73.9	673	1 LOXS_MOUSE	P48999 mus musculu
7	33	71.7	92	1 SPFS_EGOLI	P18837 escherichia
8	33	71.7	380	1 FURE_PHAOU	P32291 phaseolus a
9	33	71.7	490	1 MURE_DEIRA	Q9rx13 deinococcus
10	33	71.7	706	1 TRFE_HORSE	P27425 equus cabal
11	33	71.7	1085	1 ACPX_HUMAN	Q94833 homo sapien
12	32	69.6	377	1 PROB_MEIRO	Q92998 metothermus
13	32	69.6	466	1 SYN_BORBU	O51128 borrelia bu
14	32	69.6	631	1 TAC3_MOUSE	Q9fj11 mus musculu
15	32	69.6	678	1 SCOB_EMENI	Q00659 emericeila
16	32	69.6	699	1 VGLG_HSV2H	P13290 herpes simp
17	32	69.6	739	1 GUNF_CLOTH	P26224 clostridium
18	32	69.6	838	1 TAC3_HUMAN	Q9vfa5 homo sapien
19	31	67.4	76	1 HBPI_HUMAN	Q75506 homo sapien
20	31	67.4	196	1 TRPA_SYNPY	O02187 synchococc
21	31	67.4	285	1 LEC_SOVENB	P05046 glycine max
22	31	67.4	412	1 KSGD_ARATH	Q39010 arabidopsis
23	31	67.4	413	1 INTA_EGOLI	P32053 escherichia
24	31	67.4	424	1 CRIC_DICDI	Q23858 dictyosteli
25	31	67.4	472	1 KSGT_ARATH	Q96287 arabidopsis
26	31	67.4	545	1 AIRE_HUMAN	O43918 homo sapien
27	31	67.4	578	1 YTFW_HAEIN	P44038 haemophilus
28	31	67.4	703	1 CDGT_BACS2	P31746 bacillus sh
29	31	67.4	704	1 CDGT_BACOH	P27036 bacillus oh
30	31	67.4	704	1 ICA_PIG	Q29545 sus scrofa
31	31	67.4	708	1 TRFL_CAMDR	Q9tum0 camelus dro
32	31	67.4	719	1 HS9A_HORSE	Q9gkx7 equus cabal
33	31	67.4	720	1 DNLJ_AQUAE	O66880 aquifex aeo

ALIGNMENTS

```

RESULT 1
SYM_SYNY3
ID SYM_SYNY3 STANDARD; PRT; 532 AA.
AC Q55729;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase)
DE (MetRS).
DE
GN METG OR METS OR SLR0649.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugitara M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome."
RL DNA Res. 2:153-166(1995).
CC -!- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING
CC REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO
CC FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR
CC TRNA(EMET) AMINOACYLATION.
CC -!- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) -> AMP +
CC T9NA(EMET) AMINOACYLATION.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE
CC STRONG, TO CYSTEINYL-TRNA SYNTHETASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D64002; BAA10371.1;
CC HSSP; P23395; 1ABH.
CC InterPro: IPR002300; tRNA-synt_1a.
CC InterPro: IPR001412; tRNA-synt_1.
CC InterPro: IPR002304; tRNA-synt_met.
CC Pfam: PF001133; tRNA-synt_1.1.
CC PRINTS: PR01041; TRNASYNTHMET.
CC PROSITE: PS00178; AA-TRNA_LIGASE_1; FALSE_NEG.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 14 26 "HIGH" REGION.
CC SITE 305 309 "KMSKS" REGION.
CC BINDING 308 308 ATP (BY SIMILARITY).
CC SEQUENCE 532 AA; 61442 MW; 8D97AE90D3F18A34 CRC64;

```

P08238 homo sapien
P11499 mus musculu
O57521 brachydanio
Q04619 gallus gall
Q90474 brachydanio
P11501 gallus gall
P07900 homo sapien
P46633 cricetus
P07901 mus musculu
O02705 sus scrofa
Q9v5e3 homo sapien
Q59119 aeromonas h

34 31 67.4 723 1 HS9B_HUMAN
35 31 67.4 723 1 HS9B_MOUSE
36 31 67.4 725 1 HS9B_BRARE
37 31 67.4 725 1 HS9B_CHICK
38 31 67.4 726 1 HS9A_BRARE
39 31 67.4 728 1 HS9A_CHICK
40 31 67.4 731 1 HS9A_HUMAN
41 31 67.4 732 1 HS9A_CRIGR
42 31 67.4 732 1 HS9A_MOUSE
43 31 67.4 732 1 HS9A_PIG
44 31 67.4 794 1 CDB6_HUMAN
45 31 67.4 843 1 CYAA_AERHY

Downloaded from ascelibrary.org by University of California, San Diego on 06/01/15. Copyright ASCE, For All Rights Reserved, No part of this document may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage or retrieval system, without permission in writing from ASCE.

DR EMBL: BC009642; AA09642.1; -
DR EMBL: AF177940; AAC09301.1; -
DR InterPro: IPR000272; ATPIC1_PLM_MAT8.
DR Pfam: PF02038; ATPIC1_PLM_MAT8; 1.
DR PROSITE: PS01310; FYXD; 1.
KW Transmembrane; Signal; Ionic channel; Ion transport;
KW Alternative splicing; Polymorphism.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 178 FYXD DOMAIN-CONTAINING ION TRANSPORT
FT 21
FT DOMAIN 22 145 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 146 164 POTENTIAL.
FT DOMAIN 165 178 CYTOPLASMIC (POTENTIAL).
FT VARSPLIC 1 97 MSPSGRLCLLTIVGLILPTRGQTLKDTSSSSADSTIMDIQ
FT VPTIRAPDAVYKTELOPTSPPTWPADETPOPTOOLEGTD
FT GLVTDPEHSTKA -> MQLTSLNIPCELGSLPSTDL
FT ATLS (IN ISOFORM 2).
FT VARIANT 35 35 S -> A (IN DSNP:1688005).
FT /FTID-VAR_012349.
FT CONFLICT 140 141 HT -> SH (IN REF. 1).
FT SEQUENCE 178 AA; 19453 MW; 7429D5A838B75AA1 CRC64;
SQ

Query Match 73.9%; Score 34; DB 1; Length 178;
Best Local Similarity 71.4%; Pred. No. 7.6;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 TDPTQTHQ 8
DB 87 TDPTQTHQ 93
|||:|:
|||:|:

RESULT 4
LOX5_MESAU
ID LOX5_MESAU STANDARD; PRT; 672 AA.
AC P51399;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Arachidonate 5-lipoxygenase (EC 1.13.11.34) (5-lipoxygenase) (5-LO).
GN ALOX5.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10046;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-SYRIAN;
RX MEDLINE=97109716; PubMed=8951996;
RA Kitzler J.W., Eling T.E.;
RT "Cloning, sequencing and expression of a 5-lipoxygenase from Syrian hamster embryo fibroblasts."
RT Prostaglandins Leukot. Essent. Fatty Acids 55:269-277(1996).
RL -1- CATALYTIC ACTIVITY: ARACHIDONATE + O(2) - (6E,8Z,11Z,14Z)-(5S)-5-HYDROPEROXYCOSA-6,8,11,14-TETRAENOATE (THE PRODUCT IS RAPIDLY CONVERTED TO LEUKOTRIENE A4).
CC -1- COFACTOR: IRON, ALSO REQUIRES CALCIUM AND ATP FOR ACTIVITY.
CC -1- PATHWAY: THIS ENZYME CATALYSES THE FIRST TWO STEPS IN THE BIOSYNTHESIS OF LEUKOTRIENES.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE LIPOXYGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U43333; AAA85257.1; -
DR HSSP: P12530; 1LOX.

DR InterPro: IPR001024; LH2.
DR InterPro: IPR000907; Lipoxygenase.
DR Pfam: PF00305; lipoxygenase; 1.
DR Pfam: PF01477; PLAT; 1
DR PRINTS: PR00087; LIPOXYGENASE.
DR SMART: SM00308; LH2; 1.
DR PROSITE: PS00081; LIPOXYGENASE_2; 1.
DR PROSITE: PS00711; LIPOXYGENASE_1; 1.
KW Oxidoreductase; Dioxygenase; Iron; Leukotriene biosynthesis;
KW Calcium.
FT INIT_MET 0 0 BY SIMILARITY.
FT METAL 366 366 IRON (BY SIMILARITY).
FT METAL 371 371 IRON (BY SIMILARITY).
FT METAL 549 549 IRON (BY SIMILARITY).
FT METAL 672 672 IRON (BY SIMILARITY).
SQ SEQUENCE 672 AA; 77741 MW; 1F7234B5C297B30F CRC64;

Query Match 73.9%; Score 34; DB 1; Length 672;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDPTQTHQ 8
DB 295 KTDPTQTHQ 302
:|:|:|:
:

RESULT 5
LOX5_RAT
ID LOX5_RAT STANDARD; PRT; 672 AA.
AC P12527;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Arachidonate 5-lipoxygenase (EC 1.13.11.34) (5-lipoxygenase) (5-LO).
GN ALOX5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88330933; PubMed=3417684;
RA Balcarek J.M., Theisen T.W., Cook M.N., Varrichio A., Hwang S.-M., Strolzacker M.W., Crooke S.T.;
RT "Isolation and characterization of a cDNA clone encoding rat 5-lipoxygenase."
RL J. Biol. Chem. 263:13937-13941(1988).
RN [2]
RP REVISIONS TO 666-669.
RX MEDLINE=93298753; PubMed=8518276;
RA Minor W., Steczko J., Bolin J.T., Otwinowski Z., Axelrod B.;
RT "Crystallographic determination of the active site iron and its ligands in soybean lipoxygenase L-1."
RL Biochemistry 32:6320-6323(1993).
CC -1- CATALYTIC ACTIVITY: ARACHIDONATE + O(2) - (6E,8Z,11Z,14Z)-(5S)-5-HYDROPEROXYCOSA-6,8,11,14-TETRAENOATE (THE PRODUCT IS RAPIDLY CONVERTED TO LEUKOTRIENE A4).
CC -1- COFACTOR: IRON, ALSO REQUIRES CALCIUM AND ATP FOR ACTIVITY.
CC -1- PATHWAY: THIS ENZYME CATALYSES THE FIRST TWO STEPS IN THE BIOSYNTHESIS OF LEUKOTRIENES.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE LIPOXYGENASE FAMILY.
CC -1- CAUTION: REF.2 SUGGESTS THAT THIS PROTEIN HAS A DIFFERENT C-TERMINUS DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----


```

DR MCD: MGI:877999; Alox5.
DR InterPro: IPR001024; LH2.
DR DR InterPro: IPR000907; Lipoxxygenase.
DR Pfam: PF00305; lipoxxygenase; 1.
DR DR Pfam: PF01477; PLAT; 1.
DR PRINTS: PR00087; LIPOXYGENASE.
DR SMART: SM00308; LH2; 1.
DR PROSITE: PS00081; LIPOXYGENASE_2; 1.
DR PROSITE: PS00711; LIPOXYGENASE_1; 1.
DR KW oxidoreductase; Dioxxygenase; Iron; Leukotriene biosynthesis:
KW Calcium.
KW INIT_MET 0 0 BY SIMILARITY.
FT METAL 367 367 IRON (BY SIMILARITY).
FT FT METAL 372 372 IRON (BY SIMILARITY).
FT FT METAL 550 550 IRON (BY SIMILARITY).
FT FT METAL 673 673 IRON (BY SIMILARITY).
FT FT METAL 673 673 IRON (BY SIMILARITY).
SQ SEQUENCE 673 AA: 77068 MW: 052910477A3B7085 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 673;
Best Local Similarity 75.0%; Pred. No. 33;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps

QY 1 ETDPOTHQ 8
DB 296 KDPCTCHQ 303
:|||||
:|||||

RESULT 7
FSFB_ECOLI
ID FSFB_ECOLI STANDARD; PRT; 92 AA.
AC P18837;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Sugar fermentation stimulation protein B (Ner-like protein).
GN FSFB OR NLP OR SF57 OR B3188 OR Z4551 OR ECS4067.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]_TaxID=562, 83334;
RP SEQUENCE FROM N.A.
RX MEDLINE=89359178; PubMed=2670911;
RA Choi Y.-L., Nishida T., Kawamukai M., Utsumi R., Sakai H., Komano T.;
RT "Cloning and sequencing of an Escherichia coli gene, nlp, highly
KT homologous to the ner genes of bacteriophages Mu and D108.";
KL J. bacteriol. 171:5222-5225(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postaj G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grothbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Iken G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RN [4]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN POSITIVE REGULATION OF THE
CC METABOLISM OF SUGARS.
CC -!- SIMILARITY: BELONGS TO THE NER FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X68873; CA48736.1; .
CC EMBL: U18997; AAA57989.1; .
CC DR EMBL: AE000399; AAC76220.1; .
CC DR EMBL: AE005547; AAG58322.1; .
CC DR EMBL: AP002564; BAB37490.1; .
CC PIR: JVO027; BVBCNP.
CC HSP: P06020; INEO.
CC EcoGene: EG10656; nlp.
KW Transcription regulation; DNA-binding; Activator; Complete proteome.
FT DNA_BIND 50 69 H-T-H MOTIF (BY SIMILARITY).
FT CONFLICT 74 74 "T -> P (IN REF. 4)."
SQ SEQUENCE 92 AA; 10495 MW; 4B3859604F4259EC CRC64;

Query Match 71.7%; Score 33; DB 1; Length 92;
Best Local Similarity 83.3%; Pred. NO. 5.8;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DPQTHQ 8
| | | | |
DB 71 DPQTHE 76

RESULT 8
FD3E_PHAU STANDARD; PRT; 380 AA.
AC P32291;
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 01-OCT-1993 (Rel. 27, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Omega-3 fatty acid desaturase, endoplasmic reticulum (EC 1.14.99.-)
DE (Indole-3-acetic acid induced protein ARG1).
GN ARG1.
OS Phaseolus aureus (Mung bean) (Vigna radiata).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
OC NCBI_TaxID=3916;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hypocotyl;
RA Yamamoto K.T., Mori H., Imaseki H.;
RT "Novel mRNA sequences induced by indole-3-acetic acid in sections of
RT elongating hypocotyls of mung bean (Vigna radiata).";
RL Plant Cell Physiol 33:13-20(1992).
CC -!- FUNCTION: MICROSOAL (ER) OMEGA-3 FATTY ACID DESATURASE INTRODUCES
CC THE THIRD DOUBLED IN THE BIOSYNTHESIS OF 18:3 FATTY ACIDS,
CC IMPORTANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT TO USE
CC CYTOCHROME B5 AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS
CC ESTERIFIED TO PHOSPHATIDYLCHOLINE AND, POSSIBLY, OTHER
CC PHOSPHOLIPIDS.

```

```

CC -!- PATHWAY: POLYUNSATURATED FATTY ACID BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum.
CC -!- INDUCTION: BY AUXIN, ETHYLENE AND WOUNDING.
CC -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC AND/ OR BE INVOLVED IN METAL ION BINDING.
CC -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D14410; BAA03306.1; .
CC InterPro: IPR001225; FA_desaturase.
CC Pfam: PF00487; FA_desaturase; 1.
CC ProDom: PD001081; FA_desaturase; 1.
KW Oxidoreductase; Fatty acid biosynthesis; Endoplasmic reticulum;
KW Transmembrane.
FT TRANSMEM 59 78 POTENTIAL.
FT TRANSMEM 208 231 POTENTIAL.
FT TRANSMEM 238 256 POTENTIAL.
FT DOMAIN 97 101 HISTIDINE BOX 1.
FT DOMAIN 133 137 HISTIDINE BOX 2.
FT DOMAIN 300 304 HISTIDINE BOX 3.
SQ SEQUENCE 380 AA; 43996 MW; 1C0D5117A8DAE16B CRC64;

Query Match 71.7%; Score 33; DB 1; Length 380;
Best Local Similarity 62.5%; Pred. NO. 27;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDPTQHQ 8
| | | | |
DB 366 QTDPLHLQ 373

RESULT 9
MURE_DEIRA STANDARD; PRT; 490 AA.
AC Q9RLJ3;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-N-acetylmuramoylalanine-D-glutamate--2,6-diaminopimelate ligase
DE (EC 6.3.2.13) (UDP-N-acetylmuramyl-tripeptide synthetase) (Meso-
DE diaminopimelate-adding enzyme) (UDP-MurNAc-tripeptide synthetase).
GN MURE OR DR0297
OS Deinococcus radiodurans.
OC Bacteria; Thermus/deinococcus group; Deinococcales; Deinococcus.
OC NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Morfat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
CC -!- FUNCTION: CELL WALL FORMATION. DIAMINOPIMELIC ACID ADDING ENZYME
CC (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramyl-L-alanyl-D-
CC glutamate + meso-2,6-diaminoheptanedioate = ADP + phosphate +
CC UDP-N-acetylmuramyl-L-alanyl-D-glutamyl-meso-2,6-
CC diaminohexanedioate.

```

CC -|- PATHWAY: PEPTIDOLYCAN BIOSYNTHESIS.
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -|- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE001890; AAF09877.1; -
 CC HSP: P11880; IGG4.
 CC TIGR: DR0237; -
 CC InterPro: IPR000713; Mur_ligase.
 CC DR InterPro: IPR004101; Mur_ligase_C.
 CC DR Pfam: PF01225; Mur_ligase; 1.
 CC DR Pfam: PF02875; Mur_ligase_C; 1.
 CC DR Peptidoglycan synthetase; Cell wall; Cell division; Ligase;
 CC ATP-binding; Complete proteome.
 CC FT NP_BIND 113 119 ATP (POTENTIAL).
 CC SQ SEQUENCE 490 AA: 52316 MW: 46C748E9768754CA CRC64;
 CC
 CC Query Match 71.7%; Score 33; DB 1; Length 490;
 CC Best Local Similarity 100.0%; Pred. No. 36;
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 1 ETDPTQ 6
 CC |||||
 CC Db 16 ETDPTQ 21
 CC
 CC RESULT 10
 CC TRFE_HORSE
 CC ID TRFE_HORSE STANDARD; PRT: 706 AA.
 CC AC P27425;
 CC DT 01-AUG-1992 (Rel. 23, Created)
 CC DT 01-AUG-1992 (Rel. 23, Last sequence update)
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
 CC DE Serotransferrin precursor (Siderophilin) (Beta-1-metal binding
 CC globulin).
 CC TF.
 CC GN Equus caballus (Horse).
 CC OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC CC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 CC OX NCBI_TaxID=9796;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=93277958; PubMed=8504171;
 CC RA Carpenter M.A., Broad T.B.;
 CC RT "The cDNA sequence of horse transferrin";
 CC RL Biochim. Biophys. Acta 1173:230-232(1993).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE=Extraembryonic tissue;
 CC RA McDowell K.J., Adams M.H., Baker C.B.;
 CC RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC CC -|- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE. IT IS RESPONSIBLE FOR THE
 CC TRANSPORT OF IRON FROM SITES OF ABSORPTION AND HEME DEGRADATION
 CC TO THOSE OF STORAGE AND UTILIZATION. SERUM TRANSFERRIN MAY ALSO
 CC HAVE A FURTHER ROLE IN STIMULATING CELL PROLIFERATION.
 CC CC -|- SUBUNIT: MONOMER.
 CC CC -|- SUBCELLULAR LOCATION: Secreted.
 CC CC -|- TISSUE SPECIFICITY: PLASMA.
 CC CC -|- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC CC -|- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M69020; AAA30958.1; -
 CC DR EMBL: U21127; AAA63684.1; -
 CC DR PIR: S33761; S33761.
 CC DR HSP: P02787; IABE.
 CC DR InterPro: IPR001156; Transferrin.
 CC DR Pfam: PF00405; transferrin; 2.
 CC DR PRINTS: PR00422; TRANSFERRIN.
 CC DR SMART: SM00094; TR_FER; 2.
 CC DR PROSITE: PS00205; TRANSFERRIN_1; 2.
 CC DR PROSITE: PS00206; TRANSFERRIN_2; 2.
 CC DR PROSITE: PS00207; TRANSFERRIN_3; 2.
 CC KW Iron transport; Glycoprotein; Metal-binding; Repeat; Signal.
 CC FT SIGNAL 1 19
 CC FT CHAIN 20 706
 CC FT REPEAT 20 357
 CC FT REPEAT 358 706
 CC FT DISULFID 26 64
 CC FT DISULFID 36 55
 CC FT DISULFID 134 215
 CC FT DISULFID 174 190
 CC FT DISULFID 177 198
 CC FT DISULFID 187 200
 CC FT DISULFID 248 262
 CC FT DISULFID 360 623
 CC FT DISULFID 366 398
 CC FT DISULFID 376 389
 CC FT DISULFID 423 701
 CC FT DISULFID 441 664
 CC FT DISULFID 474 550
 CC FT DISULFID 498 632
 CC FT DISULFID 508 522
 CC FT DISULFID 519 533
 CC FT DISULFID 590 604
 CC FT DISULFID 642 647
 CC FT METAL 79 79
 CC FT METAL 111 111
 CC FT METAL 209 209
 CC FT METAL 270 270
 CC FT METAL 413 413
 CC FT METAL 449 449
 CC FT METAL 544 544
 CC FT METAL 612 612
 CC FT BINDING 140 140
 CC FT BINDING 480 480
 CC FT CARBOHYD 515 515
 CC SQ SEQUENCE 706 AA: 78094 MW: 1A0FA366C0409D8A CRC64;
 CC
 CC Query Match 71.7%; Score 33; DB 1; Length 706;
 CC Best Local Similarity 71.4%; Pred. No. 54;
 CC Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 1 ETDPTQ 7
 CC Db 104 KTEPTQ 110
 CC
 CC RESULT 11
 CC ACFX_HUMAN STANDARD; PRT: 1065 AA.
 CC ID ACFX_HUMAN
 CC AC 094833;
 CC DT 16-OCT-2001 (Rel. 40, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Transcullin-beta (Fragment).
 CC GN KIAA0728
 CC OS Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC [1]
 RN NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99087487; PubMed=9872452;
 RA Nagase T., Ishikawa K.-f., Suyama M., Kikuno R., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 5:277-286(1998).
 CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC -!- SIMILARITY: CONTAINS 7 SPECTRIN REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AB018271; BAA34448.1;
 DR HSPL: P02631; LOMD.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR003108; GAS2.
 DR InterPro: IPR002017; Spectrin.
 DR Pfam: PF00036; efhand; 2.
 DR Pfam: PF02187; GAS2; 1.
 DR Pfam: PF00435; spectrin; 7.
 DR SMART: SM00054; Eph; 2.
 DR SMART: SM00243; GAS2; 1.
 DR SMART: SM00150; SPEC; 5.
 DR PROSITE: PS00018; EF-HAND; 2.
 DR Actin-binding; Cytoskeleton; Calcium-binding; Repeat.
 FT NON_TER 1 1
 FT REPEAT 1 74 SPECTRIN 1.
 FT REPEAT 77 183 SPECTRIN 2.
 FT REPEAT 186 293 SPECTRIN 3.
 FT REPEAT 300 401 SPECTRIN 4.
 FT REPEAT 404 510 SPECTRIN 5.
 FT REPEAT 522 578 SPECTRIN 6.
 FT REPEAT 626 650 SPECTRIN 7.
 FT CA_BIND 692 703 EF-HAND 1 (POTENTIAL).
 FT CA_BIND 728 739 EF-HAND 2 (POTENTIAL).
 SQ SEQUENCE 1065 AA; 121325 MW; 1701547c12d4249e CRC64;
 Query Match 71.7%; Score 33; DB 1; Length 1065;
 Best Local Similarity 75.0%; Pred. No. 84;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ETDPTQTHQ 8
 DB 1023 ETVPTQTHR 1030
 RESULT 12
 PROB_MEIRU STANDARD; PRT; 377 AA.
 AC Q9ZG58;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Glutamate 5-kinase (EC 2.7.2.11) (Gamma-glutamyl kinase) (GK).
 GN PROB.
 OS Meiothermus ruber.
 OC Bacteria; Thermus/Deinococcus group; Thermus group; Meiothermus.
 OX NCBI_TaxID=277;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-40;
 RA Yaklichkin S.Y., Zimina M.S., Yurchenko Y.V., Hromov I.S.,
 RA Neumivakin L.V.;
 RT "Molecular cloning and sequence analysis of the proA gene from
 RT thermophilic eubacterium Thermus ruber.";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Catalyzes the transfer of a phosphate group to glutamate
 CC to form glutamate 5-phosphate which rapidly cyclizes to 5-
 CC oxoproline.
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate = ADP + L-glutamate 5-
 CC phosphate.
 CC -!- PATHWAY: Proline biosynthesis pathway; first step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE GLUTAMATE 5-KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF082661; AAC72812.1; ALT_INIT.
 DR InterPro: IPR001048; Aakkinase.
 DR InterPro: IPR001057; Glut_5_kinase.
 DR InterPro: IPR002478; PUA.
 DR Pfam: PF00696; aakkinase; 1.
 DR Pfam: PF01472; PUA; 1.
 DR PRINTS: PR00474; GLU5KINASE.
 DR SMART: SM00359; PUA; 1.
 DR PROSITE: PS00902; GLUTAMATE_5_KINASE; 1.
 DR Transference; Kinase; Proline biosynthesis.
 KW TRANSFERASE 377 AA; 40591 MW; F637DFE9D81155 CRC64;
 SQ SEQUENCE 377 AA; 40591 MW; F637DFE9D81155 CRC64;
 Query Match 69.6%; Score 32; DB 1; Length 377;
 Best Local Similarity 71.4%; Pred. No. 43;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ETDPTQH 7
 DB 181 EADPTRH 187
 RESULT 13
 SYN_HORBU STANDARD; PRT; 466 AA.
 ID 051128;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Asparaginyl-tRNA synthetase (EC 6.1.1.22) (Asparagine--tRNA ligase)
 DE (AsnRS).
 GN ASNS OR BB0101.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Uitterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 RT burgdorferi.";
 RL Nature 390:580-586(1997).

CC -!- CATALYTIC ACTIVITY: ATP + L-asparagine + LRNA(Asn) = AMP +
CC diphosphate + L-asparagyl-LRNA(Asn).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE001123; AAC66501.1; -
CC HSSP: Q52428; 188A.
CC TIGR: B80101; -
CC InterPro: IPR002106; AA_TRNA_Ligase_IL1.
CC InterPro: IPR002309; tRNA-synt_2.
CC InterPro: IPR002312; tRNA-synt_2; 1.
CC Pfam: PF00152; tRNA-synt_2; 1.
CC Pfam: PF01336; tRNA_ant1; 1.
CC PRINTS: PR01042; TRNASYNTHASP.
CC PROSITE: PS00179; AA_TRNA_LIGASE_IL1; 1.
CC PROSITE: PS00339; AA_TRNA_LIGASE_IL2; FALSE_NEG.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC KW
CC SEQUENCE 456 AA; 53600 MW; 2B5900C0E6BDBDf6 CRC64;
CC
CC Query Match 69.6%; Score 32; DB 1; Length 466;
CC Best Local Similarity 71.4%; Pred. No. 54;
CC Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 1 ETDPTH 7
CC | | | | |
CC DB 104 ETDPTY 110
CC
CC RESULT 14
CC TAC3_MOUSE
CC ID TAC3_MOUSE STANDARD; PRT; 631 AA.
CC AC Q9JUL1; Q9WVK9;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Transforming acidic coiled-coil containing protein 3 (ARNT
CC interacting protein).
CC GN TACC3 OR AINT.
CC OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC NCBI_TaxID=10090;
CC [1]
CC SEQUENCE FROM N.A. (ISOFORM 1).
CC STRAIN-NIH Swiss; TISSUE=Embryo;
CC MEDLINE=20480355; PubMed=11025203;
CC RA Sadek C.M., Jalaguer S., Feeney E.P., Aitola M., Damdimopoulos A.E.,
CC Peltto-Huikko M., Gustafsson J.-A.;
CC "Isolation and characterization of AINT: a novel ARNT interacting
CC protein expressed during murine embryonic development.";
CC Mech. Dev. 97:13-26(2000).
CC [2]
CC SEQUENCE FROM N.A. (ISOFORM 2).
CC MEDLINE=9296831; PubMed=10366448;
CC RA Still I.H., Vince P., Cowell J.K.;
CC "The third member of the transforming acidic coiled coil-containing
CC gene family, TACC3, maps in 4p16, close to translocation breakpoints
CC in multiple myeloma, and is upregulated in various cancer cell
CC lines";
CC Genomics 58:165-170(1999).
CC -!- FUNCTION: MAY BE INVOLVED IN THE CONTROL OF CELL GROWTH AND
CC DIFFERENTIATION. MAY HAVE A ROLE IN EMBRYONIC DEVELOPMENT.

CC -!- SUBUNIT: THE COILED COIL C-TERMINUS REGION INTERACTS WITH AH
CC RECEPTOR NUCLEAR TRANSLATOR PROTEIN (ARNT) AND ARNT2.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EMBRYONICALLY EXPRESSED.
CC -!- DEVELOPMENTAL STAGE: AT 9 DAYS POSTCOITUM (PC), THE EXPRESSION IS
CC STRONG IN THE NEUROEPITHELIUM OF NEURAL TUBE AND IN PLACENTA. AT
CC 13 DAYS PC, THE EXPRESSION IS STILL OBSERVED IN NEUROEPITHELIUM.
CC FURTHERMORE, STRONG EXPRESSION IS SEEN IN LUNG, KIDNEY,
CC INTESTINES, THYMUS AND LIVER, AND A MODERATE SIGNAL IS DETECTED IN
CC THE CARTILAGE PRIMORDIUM OF DEVELOPING RIBS, TOOTH AND EYE. BY 17
CC DAYS PC, THE TISSUE DISTRIBUTION CHANGES SO THAT NO SIGNAL IS
CC DETECTED IN THE LIVER AND THE SIGNAL HAS DIMINISHED IN OTHER
CC ORGANS. IT IS OBSERVED FOR THE FIRST TIME IN THE SALIVARY GLAND,
CC THYROID GLAND AND BROWN FAT AND WAS STRONG IN THE THYMUS, EYE,
CC OLFACTORY EPITHELIUM AND CENTRAL NERVOUS SYSTEM. AT 1.5 DAYS AFTER
CC BIRTH, THE EXPRESSION IS STILL STRONG IN THYMUS, BUT WEAKER AND
CC MORE LIMITED IN BRAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF156934; AAF85763.1; -
CC EMBL: AF093542; AAD25963.1; -
CC MGD: MGI:1341163; Tacc3.
CC Coiled coil; Alternative splicing.
CC DOMAIN 431 630 COILED COIL (POTENTIAL).
CC VARSPLIC 182 277 MISSING (IN ISOFORM 2).
CC FT VARSPLIC 337 337 S -> SSSLNSQ (IN ISOFORM 2).
CC FT CONFLICT 171 171 P -> S (IN REF. 2).
CC FT CONFLICT 407 411 TPVWS -> PLCV (IN REF. 2).
CC SEQUENCE 631 AA; 70626 MW; 92D6324D3890E9C8 CRC64;
CC
CC Query Match 69.6%; Score 32; DB 1; Length 631;
CC Best Local Similarity 83.3%; Pred. No. 75;
CC Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 3 DPQTHQ 8
CC | | | | |
CC DB 63 DPQTHR 68
CC
CC RESULT 15
CC SCOB_EMBL
CC ID SCOB_EMBL STANDARD; PRT; 678 AA.
CC AC Q00659;
CC DT 15-JUL-1998 (Rel. 36, Created)
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Sulfur metabolite repression control protein.
CC GN SCOB OR MAP81.
CC OS Emericella nidulans (Aspergillus nidulans).
CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
CC Eurotiales; Trichocomaceae; Emericella.
CC NCBI_TaxID=5072;
CC [1]
CC SEQUENCE FROM N.A.
CC RA Natorff R.;
CC RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REGULATORY PROTEIN THAT CONTROL SULFUR METABOLITE
CC REPRESSION.
CC -!- SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: CONTAINS 1 F-BOX DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MET30/SCOB/SCON-2 FAMILY OF WD-REPEAT
CC PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: U21220; AAC15905.1; -
 DR HSP: P04901; LGG2.
 DR InterPro: IPR001810; F-box.
 DR InterPro: IPR001680; WPA0.
 DR Pfam: PF00646; F-box; 1.
 DR Pfam: PF00400; WD40; 7.
 DR PRINTS: PR00320; GPROTEINBKPT.
 DR SMART: SM00256; FBOX; 1.
 DR SMART: SM00320; WD40; 7.
 DR PROSITE: PS00181; FBOX; 1.
 DR PROSITE: PS00678; WD_REPEATS_1; 4.
 DR PROSITE: PS00082; WD_REPEATS_2; 7.
 DR PROSITE: PS0294; WD_REPEATS_REGION; 1.
 KW Transcription regulation; Repeat; WD repeat.
 FT DOMAIN 178 224 F-BOX.
 FT REPEAT 347 375 WD 1.
 FT REPEAT 387 415 WD 2.
 FT REPEAT 427 455 WD 3.
 FT REPEAT 466 496 WD 4.
 FT REPEAT 508 543 WD 5.
 FT REPEAT 553 595 WD 6.
 FT REPEAT 607 635 WD 7.
 FT REPEAT 647 675 WD 8.
 SQ SEQUENCE 678 AA: 76070 MW: D840D452E37B4C53 CRC64:

Query Match 69.6%; Score 32; DB 1; Length 678;
 Best Local Similarity 71.4%; Pred. No. 81;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ETDPOTH 7
 DB 39 ETEPDTH 45

Search completed: August 7, 2002, 06:47:37
 Job time: 209 sec

Genfore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 06:44:33 : Search time 24.98 Seconds
(without alignments)
55.403 Million cell updates/sec

Title: US-10-020-139-2_COPY_159_166

Perfect score: 46

Sequence: 1 ETDPQTHQ 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17299429 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	100.0	249	4 Q9BQ00	Q9BQ00 homo sapien
2	46	100.0	249	4 Q96DR5	Q96DR5 homo sapien
3	39	84.8	438	10 Q9LEGO	Q9LEGO lycopersico
4	38	82.6	262	12 Q86975	Q86975 phocid herp
5	37	80.4	445	10 Q22470	Q22470 oryza sativ
6	37	80.4	445	10 Q22471	Q22471 oryza sativ
7	36	78.3	880	10 Q9SVU3	Q9SVU3 arabidopsis
8	36	78.3	898	10 Q9SSB6	Q9SSB6 arabidopsis
9	36	78.3	924	15 Q83393	Q83393 mouse mamma
10	36	78.3	924	10 Q949M4	Q949M4 arabidopsis
11	35	76.1	250	10 Q9STH6	Q9STH6 arabidopsis
12	35	76.1	286	10 Q9SPG1	Q9SPG1 arabidopsis
13	35	76.1	338	13 Q9PTJ2	Q9PTJ2 brachydanio
14	35	76.1	425	16 Q9PAU9	Q9PAU9 xylella fas
15	35	76.1	444	10 Q22402	Q22402 nicotiana t
16	35	76.1	867	10 Q9LV72	Q9LV72 arabidopsis

ALIGNMENTS

RESULT 1

Q9BQ00 Q9BQ00 PRELIMINARY: PRT; 249 AA.
AC Q9BQ00:
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE BA49G10.1 (SIMILAR TO BOVINE SALIVARY PROTEIN BSP30).
GN BA49G10.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tracey A.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL121901; CAC03546.1; -
SQ SEQUENCE 249 AA; 27011 MW; E64E0794A1B4DB7D CRC64;

Query Match 100.0%; Score 46; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. NO. 0.21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDPQTHQ 8

Db 159 ETDPQTHQ 166

RESULT 2

Q96DR5 Q96DR5 PRELIMINARY: PRT; 249 AA.
AC Q96DR5:
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE PAROTID SECRETORY PROTEIN.
OS Homo sapiens (Human).

Q9P039 homo sapien
Q96db9 homo sapien
Q24532 volvox cart
Q969w4 homo sapien
Q9ut15 homo sapien
Q99483 homo sapien
Q00029 acromonium
Q9rtx0 streptomyce
Q9bap9 homo sapien
Q98k87 rhizobium l
Q97e2 chlamydia p
Q9jsa5 chlamydia p
Q9ft45 arabidopsis
Q65744 cicet ariet
Q9s164 arabidopsis
Q91398 spingomona
Q9bc91 homo sapien
Q931y4 pravotella
Q9jh67 sheep astro
Q9w554 drosophila
Q9h555 homo sapien
Q91v14 mus musculu
Q27416 methanother
Q9p2a8 homo sapien
Q91zu7 mus musculu
Q91zu6 mus musculu
Q9q0d8 herpes simp
Q9q0d7 herpes simp
Q9q0d6 herpes simp

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PAROTID;
 RA Venkatesh S.G.; Geetha C.; Gorr S.-U.;
 RT "A member of the PSP/plunc family of BP1 proteins is expressed in the
 RT human parotid gland."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF432917; AAL28113.1; -
 SQ SEQUENCE 249 AA; 27110 MW; FD54B624A1A4CA7C CRC64;

Query Match 100.0%; Score 46; DB 4; Length 249;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDPOTHQ 8
 |||||
 Db 159 ETDPOTHQ 166

RESULT 3
 Q9LEGO
 ID O9LEGO PRELIMINARY; PRT; 438 AA.
 AC O9LEGO;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE GDP DISSOCIATION INHIBITOR (FRAGMENT).
 GN GDI1.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asceridae; easterids I; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bary R.; Chmelitsky I.; Sobolev I.; Salts Y.;
 RT "A tomato GDP dissociation inhibitor gene highly expressed in
 RT petals."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ401079; CAB94202.1; -
 DR HSSP; P21856; IGND.
 DR InterPro: IPR002005; Rab_GDI_REP.
 DR Pfam: PF00996; GDI; 1.
 DR PRINTS; PR00891; KAUGHIRKP.
 FT NON_TER 1
 SQ SEQUENCE 438 AA; 48728 MW; R51E7E02D956DC2A5 CRC64;

Query Match 84.8%; Score 39; DB 10; Length 438;
 Best Local Similarity 75.0%; Pred. No. 9.3;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDPOTHQ 8
 |||||
 Db 144 ETDPOTHQ 151

RESULT 4
 Q86975
 ID Q86975 PRELIMINARY; PRT; 262 AA.
 AC Q86975;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE U152; HELICASE-PRIMASE COMPLEX HOMOLOG PROTEIN (FRAGMENT).
 OS phocid herpesvirus 2.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.
 OX NCBI_TaxID=47419;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=7848;
 RX MEDLINE=96145132; PubMed=8558126;
 RA Harder T.C.; Harder M.; Vos H.; Kulonen K.; Kennedy-Stoskopf S.;
 KA Liess B.; Appel M.J.; Osterhaus A.D.;
 RT "Characterization of phocid herpesvirus-1 and -2 as putative alpha-and
 RT gammaherpesviruses of North American and European pinnipeds."
 RL J. Gen. Virol. 77:27-35(1996).
 DR EMBL: S81230; AAB35973.1; -
 KW Helicase.
 FT NON_TER 1
 SQ SEQUENCE 262 AA; 30101 MW; 83AA01A8BB6D2790 CRC64;

Query Match 82.6%; Score 38; DB 12; Length 262;
 Best Local Similarity 75.0%; Pred. No. 8.6;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDPOTHQ 8
 |||||
 Db 163 ETDPOTHQ 170

RESULT 5
 Q22470
 ID Q22470 PRELIMINARY; PRT; 445 AA.
 AC Q22470;
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE GDP DISSOCIATION INHIBITOR PROTEIN OSGD11.
 GN OSGD11.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kim W.Y.; Kim C.Y.; Cheong N.E.; Choi Y.O.; Lee K.O.; Lee S.H.;
 RA Park J.B.; Nakano A.; Bahk J.D.; Cho M.J.; Lee S.Y.;
 RT "Characterization of two fungal-elicitor-induced rice cDNAs encoding
 RT functional homologues of the rab-specific GDP-dissociation
 RT inhibitor."
 RL Planta 210:143-149(1999).
 DR EMBL: AF016896; AAB69870.1; -
 DR HSSP; P21856; IGND.
 DR InterPro: IPR002005; Rab_GDI_REP.
 DR Pfam: PF00996; GDI; 1.
 DR PRINTS; PR00891; RABGDIREP.
 SQ SEQUENCE 445 AA; 49701 MW; 09EEC9101849FD3 CRC64;

Query Match 80.4%; Score 37; DB 10; Length 445;
 Best Local Similarity 75.0%; Pred. No. 24;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDPOTHQ 8
 |||||
 Db 151 EADPKTHQ 158

RESULT 6
 Q22471
 ID Q22471 PRELIMINARY; PRT; 445 AA.
 AC Q22471;
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE GDP DISSOCIATION INHIBITOR PROTEIN OSGD12.
 GN OSGD12.
 OS Oryza sativa (Rice).

1 ETDPTHQ 8

```

RESULT      9
QB3393
ID          PRELIMINARY;      PRT;      899 AA.
AC          QB3393
DT          01-NOV-1996 (TrEMBLrel. 01, Created)
DT          01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT          01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE          GAG PRO POL POLYPROTEIN (FRAGMENT).
GN          POL OR GAG-POL OR GAG PRO POL.
GS          Mouse mammary tumor virus.
SS          Viruses; Retroid viruses; Retroviridae; Betaretrovirus.
OC          NCBI_TaxID=11757;
OX          [1]
OR          SEQUENCE FROM N.A.
RP          MEDLINE=87112944; PubMed=3027377;
RR          Moore R., Dixon M., Smith R.E., Peters G., Dickson C.;
RT          "Complete nucleotide sequence of a milk-transmitted mouse mammary
RT          tumor virus: two frameshift suppression events are required for
RT          translation of gag and pol.";
RL          J. Virol. 61:480-490(1987).
RL          [2]
RP          SEQUENCE FROM N.A.
RR          PubMed=11091515;
RR          Nishito M., Xu L., Sasaki M., Haga S., Okumoto M., Mori N.,
RR          Sarkar N.H., Acha-orbea H., Enami J., Imai S.;
RR          "Complete Nucleotide Sequence of mouse mammary tumor virus from JYG
RR          chinese wild mice; absence of bacterial insertion sequences in the
RR          cloned viral gag gene.";
RR          Breast cancer 1:89-94(1994).
RR          -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS
CC          (BY SIMILARITY).
CC          EMBL; D16249; BAA03767.1; ..
CC          HSPSP; P03355; 1MMI.
OR          OR

```

DR InterPro: IPR001037; Integrase_C.
DR InterPro: IPR003308; Integrase_Zn.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVtse.
DR Pfam: PF00552; Integrase_1.
DR Pfam: PF02022; Integrase_Zn; 1.
DR Pfam: PF00075; rnaseh; 1.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rvt; 1.
KW DNA-binding; Endonuclease; Hydrolase; Nucleotidyltransferase;
KW Polypeptide; RNA-directed DNA polymerase.
FT NON_TER
SQ SEQUENCE 899 AA; 102359 MW; A78CAF65FF2F0C34 CRC64;

Query Match 78.3%; Score 36; DR 15; Length 899;
Best Local Similarity 75.0%; Pred. No. 77;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 ETDPTQHQ 8
|||||
DB 873 ETDPRPHQ 890

RESULT 10
Q949M4 PRELIMINARY; PRT; 924 AA.
ID Q949M4
AC Q949M4
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 102.9 KDA PROTEIN.
GN AT4G28760.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
RA Carninci P., Chen H., Chen R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.,
RT "Full length cDNA of gene AT4G28760 (GI:7269731)."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY051010; AAK93687.1;
KW Hypothetical protein.
SQ SEQUENCE 924 AA; 102903 MW; 10B55CC87876CA86 CRC64;

Query Match 78.3%; Score 36; DR 10; Length 924;
Best Local Similarity 87.5%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 ETDPTQHQ 8
|||||
DB 120 ETLPTQHQ 127

RESULT 11
Q9STH6 PRELIMINARY; PRT; 250 AA.
ID Q9STH6
AC Q9STH6
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PUTATIVE TRANSCRIPTION FACTOR.

GN T4C9.190 OR AT4G12350.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Peters S.A., van Staveren M., Dirkse W., Stiekema W.,
RA Bancroft J., Mewes H.W., Mayer K.F.X., Lemcke K., Mannhaupt C.,
RA Schueller C.,
KL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Peters S.A., van Staveren M., Dirkse W., Stiekema W., Mewes H.W.,
RA Lemcke K., Mayer K.F.X.,
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL080318; CAB45982.1;
DR EMBL: AL161533; CAB78278.1;
DR HSSP: Q03237; IASJ.
DR InterPro: IPR001005; Myb_DNA_bind.
DR Pfam: PF00249; myb_DNA-binding; 2.
DR SMART: SM00395; SANT; 1.
DR PROSITE: PS00334; MYB_2; 1.
DR PROSITE: PS00090; MYB_3; 1.
SQ SEQUENCE 250 AA; 28231 MW; D97D7A56A13D8A89 CRC64;

Query Match 76.1%; Score 35; DB 10; Length 250;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 1 ETDPTQHQ 8
|||||
DB 84 ETDPTQHQ 91

RESULT 12
Q9SPG1 PRELIMINARY; PRT; 286 AA.
ID Q9SPG1
AC Q9SPG1
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE TRANSCRIPTION FACTOR.
GN MYB42.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21481677; PubMed=11597504;
RA Stracke R., Weiber M., Weisshaar B.,
RT "The R2R3-MYB gene family in Arabidopsis thaliana."
RL Curr. Opin. Plant Biol. 4:447-456(2001).
DR EMBL: AF175999; AAD53104.1;
DR HSSP: P01103; IPOM.
DR InterPro: IPR001005; Myb_DNA_bind.
DR Pfam: PF00249; myb_DNA-binding; 2.
DR SMART: SM00395; SANT; 2.
DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
DR PROSITE: PS00334; MYB_2; 1.
DR PROSITE: PS00090; MYB_3; 2.

SQ SEQUENCE 286 AA: 32402 MW: A266805A0E784A61 CRC64;
 Query Match 76.1%; Score 35; DB 10; Length 286;
 Best Local Similarity 75.0%; Pred. No. 37;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ETDPTQHQ 8
 DB 120 EIDPSTHQ 127
 RESULT 13
 Q9PTJ2 PRELIMINARY: PRT: 338 AA.
 AC Q9PTJ2;
 DT 01-MAY-2000 (TREMHLrel. 13, Created)
 DT 01-MAY-2000 (TREMHLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMHLrel. 19, Last annotation update)
 DE ZINC FINGER TRANSCRIPTION FACTOR GATA4.
 GN GATA4.
 OS Brachydanio rerio (zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Danio.
 OC NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Reiter J.F., Zon L.I., Stainier D.Y.R.;
 RT "Zebrafish gata4 and gata6 are expressed in the heart and endoderm."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF191577; AAF15275.1;
 DR HSSP: P17679; 1GNF.
 DR ZFIN: ZDB-GENE-980526-476; gata4.
 DR InterPro: IPR000679; Znf_GATA.
 DR InterPro: IPR001164; Znf_GCS.
 DR Pfam: PF00320; GATA; 2.
 DR PRINTS: PR00619; GATAZNFINGER.
 DR SMART: SM00401; Znf_GATA; 2.
 DR PROSITE: PS00344; GATA_ZN_FINGER_1; 2.
 DR PROSITE: PS01114; GATA_ZN_FINGER_2; 2.
 SQ SEQUENCE 338 AA: 36922 MW: 862F278C445AAA1F CRC64;

Query Match 76.1%; Score 35; DB 13; Length 338;
 Best Local Similarity 75.0%; Pred. No. 44;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ETDPTQHQ 8
 DB 293 EKDPDTHQ 300
 RESULT 14
 Q9PAU9 PRELIMINARY: PRT: 425 AA.
 AC Q9PAU9;
 DT 01-OCT-2000 (TREMHLrel. 15, Created)
 DT 01-OCT-2000 (TREMHLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMHLrel. 19, Last annotation update)
 DE AMINOTRANSFERASE.
 GN XF2396.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OC NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Almeida R., Alves L.M.C., Arra J.B., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,

RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincou A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohelsel J.D., Junqueira M.B., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nauai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira K.C., Palmeri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";
 RL Nature 406:151-159(2000).
 DR EMBL: AE004048; AAF85195.1;
 DR InterPro: IPR001176; ACCSYNTHASE.
 DR PRINTS: PR00753; ACCSYNTHASE.
 KW Complete proteome.
 SQ SEQUENCE 425 AA: 46381 MW: 5B1A2C657AFC39B6 CRC64;
 Query Match 76.1%; Score 35; DB 16; Length 425;
 Best Local Similarity 85.7%; Pred. No. 56;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 TDIQTHQ 8
 DB 71 TDPYTHQ 77
 RESULT 15
 O22402 PRELIMINARY: PRT: 444 AA.
 AC O22402;
 DT 01-JAN-1998 (TREMHLrel. 05, Created)
 DT 01-JAN-1998 (TREMHLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMHLrel. 19, Last annotation update)
 DE GDP DISSOCIATION INHIBITOR.
 GN GDI.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 OC NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. SAMSUN;
 RA Ezaki B., Koyanagi M., Gardner R.C., Matsumoto H.;
 RT "Nucleotide sequence of cDNA for a GDP dissociation inhibitor (GDI)
 RT which is induced by aluminum (Al) ion stress in tobacco cell culture
 RT (Accession No. AF012823) (PGR97-133).";
 RL Plant Physiol. 115:314-314(1997).
 DR EMBL: AF012823; AAB80717.1;
 DR HSSP: P21856; 1GND.
 DR InterPro: IPR002005; Rab_GDI_REP.
 DR Pfam: PF00996; GDI; 1.
 DR PRINTS: PR00891; RABGDIREP.
 SQ SEQUENCE 444 AA: 49702 MW: 5484214FAD3773D3 CRC64;
 Query Match 76.1%; Score 35; DB 10; Length 444;
 Best Local Similarity 62.5%; Pred. No. 59;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDPQTHQ 8
|:|:|:
Db 151 EDPKTHE 158

Search completed: August 7, 2002, 06:48:10
Job time: 217 sec

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 27; SEQ ID NO 34070; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 50 AA;

Query Match 100.0%; Score 46; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ETDPTQHQ 8
Db 22 etdptqhq 29
|||||

RESULT 2
AAM62308
ID AAM62308 standard; Protein; 50 AA.
XX
XX AAM62308;
XX
XX 05-NOV-2001 (first entry)

Human brain expressed single exon probe encoded protein SEQ ID NO: 34413.
Human; brain expressed exon; gene expression analysis; probe;
microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
epilepsy; cancer.
Homo sapiens.
WO200157275-A2.
09-AUG-2001.

30-JAN-2001; 2001WO-US00667.
04-FEB-2000; 2000US-0180312.
26-MAY-2000; 2000US-0207456.
30-JUN-2000; 2000US-0608408.
03-AUG-2000; 2000US-0632366.
21-SEP-2000; 2000US-0234687.
27-SEP-2000; 2000US-0236359.
04-OCT-2000; 2000GB-0024263.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI: 2001-483446/52.
Single exon nucleic acid probes for analyzing gene expression in human
brains -

Example 4; SEQ ID NO: 34413; 650pp + Sequence Listing; English.
The present invention provides a number of single exon nucleic acid
probes which are derived from genomic sequences expressed in the human
brain. They can be used to measure gene expression in brain cell samples,
which may enable the diagnosis and improved treatment of nervous system
diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
XX Sequence 50 AA;

Query Match 100.0%; Score 46; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ETDPTQHQ 8
Db 22 etdptqhq 29
|||||

RESULT 3
AAM75111
ID AAM75111 standard; Protein; 50 AA.
XX
XX AAM75111;
XX
XX 06-NOV-2001 (first entry)

Human bone marrow expressed probe encoded protein SEQ ID NO: 35417.
Human; bone marrow expressed exon; gene expression analysis; probe;
microarray; cancer; leukaemia; lymphoma; myeloma.
Homo sapiens.
WO200157276-A2.
09-AUG-2001.
30-JAN-2001; 2001WO-US00668.
04-FEB-2000; 2000US-0180312.
26-MAY-2000; 2000US-0207456.
30-JUN-2000; 2000US-0608408.
03-AUG-2000; 2000US-0632366.
21-SEP-2000; 2000US-0234687.
27-SEP-2000; 2000US-0236359.
04-OCT-2000; 2000GB-0024263.
(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;
WPI: 2001-488900/53.
Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human bone marrow -
Example 4; SEQ ID NO: 35417; 658pp + Sequence Listing; English.
The present invention provides a number of single exon nucleic acid
probes which are derived from genomic sequences expressed in the human
bone marrow. They can be used to measure gene expression in bone marrow
samples, which may enable the improved diagnosis and treatment of cancers
such as lymphoma, leukaemia and myeloma. The present sequence is a
protein encoded by one of the probes of the invention.

Query Match 100.0%; Score 46; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ETDPTQHQ 8
Db 22 etdptqhq 29
|||||

KW ulcerative colitis; Crohn's disease; atrophic gastritis.

XX Homo sapiens.

XX WO9821329-A1.

XX PD 22-MAY-1998.

XX PF 07-NOV-1997; 97WO-US20651.

XX PR 14-NOV-1996; 96US-074288.

XX PA (INCY-) INCYTE PHARM INC.

XX PI Bandman O, Goli SK;

XX WPI: 1998-297933/26.

XX DR N-PSDB; AAV37699.

XX PT New parotid secretory protein - useful for, e.g. treatment of cancer
PT and auto-immune disease, particularly of secretory or
PT gastrointestinal tissues

XX PS Claim 1: Fig 1A-C; 65pp; English.

XX CC This represents a human parotid secretory protein (HPSP). Antagonists
CC that bind specifically to, and modulate activity of HPSP are used to
CC treat cancer and autoimmune diseases particularly of secretory or
CC gastrointestinal tissue, e.g. cancer of salivary gland, thyroid,
CC prostate, breast, gastrointestinal tract or pancreas, Sjorgen's syndrome,
CC Graves disease, thyroiditis, insulin-dependent diabetes, pancreatitis,
CC ulcerative colitis, Crohn's disease and atrophic gastritis. Host cells
CC containing expression vectors comprising the HPSP nucleic acid are used
CC to produce recombinant HPSP which is used to generate antibodies and to
CC screen for its antagonists. Antibodies are useful directly as
CC antagonists, to transport drugs to HPSP-expressing cells, to detect cells
CC that express HPSP, to monitor patients being treated with HPSP, and for
CC purification of HPSP from natural sources. Expression of HPSP may
CC indicate cell proliferation. HPSP nucleic acid or its fragments are used
CC to detect HPSP-encoding sequences (optionally after amplification by PCR)
CC by hybridisation, particularly for diagnosis and monitoring of disease,
CC but also for mapping the chromosomal sequence.

XX SQ Sequence 249 AA;

Query Match 100.0%; Score 46; DB 19; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.91;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 ETDPTQHQ 8

Db 159 etdptqhq 166

RESULT 7

ID AAB24069

XX AAB24069 standard; Protein: 249 AA.

XX AC AAB24069;

XX DT 29-JAN-2001 (first entry)

XX DE Human PRO1025 protein sequence SEQ ID NO:38.

XX KW Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
KW proliferation; tumorigenesis; identification; cancer; cytostatic;
KW neotropic; neuroprotective; antiinflammatory; immunosuppressive;
KW immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
KW neuronal disorder; glial disorder; astrocytoma; angiogenic;
KW hypothalamic disorder; glandular disorder; macrophage disorder;
KW epithelial disorder; stromal disorder; blastocoele disorder;
KW inflammatory disorder; immunologic disorder.

XX OS

XX Homo sapiens.

XX WO200053755-A2.

XX PD 14-SEP-2000.

XX PF 06-JAN-2000; 2000WO-US00376.

XX PR 08-MAR-1999; 99WO-US05028.

XX PR 02-JUN-1999; 99WO-US12252.

XX PR 23-JUN-1999; 99US-0141037.

XX PR 07-JUL-1999; 99US-0143048.

XX PR 26-JUL-1999; 99US-0145698.

XX PR 30-NOV-1999; 99WO-US28313.

XX PR 05-DEC-1999; 99WO-US30911.

XX PR 05-JAN-2000; 2000WO-US00219.

XX PA (GETH) GENENTECH INC.

XX PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;

XX PI Watanabe CK, Wood WI;

XX WPI: 2000-572270/53.

XX DR N-PSDB; AAC58379.

XX PT Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
PT treatment, diagnosis and prevention of cancer.

XX PS Claim 61: Fig 26; 286pp; English.

XX CC The present invention describes an isolated antibody that binds to
CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO335,
CC PRO519, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
CC PRO1025, PRO1030, PRO1037, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
CC PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
CC growth. The PRO polypeptides and nucleotides are useful in the
CC treatment, diagnosis and prevention of cancer. The antibodies and other
CC anti-tumour compounds may be used to treat various conditions, including
CC those characterised by overexpression and/or activation of the amplified
CC PRO genes. Exemplary conditions or disorders to be treated with such
CC antibodies and other compounds include benign or malignant tumours
CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
CC leukaemias and lymphoid malignancies, and other disorders such as neuronal,
CC glial, astrocytic, hypothalamic and other glandular, macrophagal,
CC epithelial, stromal and blastocoele disorders, and inflammatory,
CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
CC primers and hybridisation probes used in the isolation of the human PRO
CC sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
CC PRO polynucleotide and protein sequences given in the exemplification of
CC the present invention.

XX SQ Sequence 249 AA;

Query Match 100.0%; Score 46; DB 21; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.91;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDPTQHQ 8

Db 159 etdptqhq 166

RESULT 8

AAB25765

ID AAB25765 standard; Protein: 249 AA.

XX AC AAB25765;

XX


```

DF 28-NGJ-2000 (first entry)
XX Human secreted protein SEQ ID #77.
DE
XX
XX Human; secreted protein; forensic procedure; gene therapy;
KW chromosome mapping; cancer; autoimmune disease; cardiovascular disorder;
KW cystic fibrosis; hypothyroidism; immunological disorder; amyloidosis;
KW brain disorder; skeletal muscle disorder; eye disorder; obesity;
KW mitochondrialcytopathy; diabetes; atherosclerosis; Alzheimer's disease;
KW neurodegenerative disorder; graft rejection; dementia; hyperlipidaemia;
KW septic shock; impotence.
XX
OS Homo sapiens.
XX WO200037491-A2.
XX
XX 29-JUN-2000.
XX
XX 20-DEC-1999; 99WO-IB02058.
XX
XX 22-DEC-1998; 98US-0113686.
XX 25-JUN-1999; 99US-0141032.
XX
XX (GEST ) GENSET.
XX
XX Bougueleret L, Dumas J, Duclert A;
XX
XX WPI; 2000-442637/38.
XX N-PSDB: AAB87727.
XX
XX Polynucleotides and polypeptides encoding proteins with signal
XX peptides, useful in diagnostic, forensic, gene therapy and chromosome
XX mapping procedures -
XX
XX Claim 9; Figure 10; 306pp; English.
XX
XX This sequence represents a human secreted protein amino acid sequence.
XX The invention relates to sequences AAB87725-A87774 which encode human
XX secreted proteins AAB25763-B25812. The proteins include signal peptides.
XX Included in the invention are a host cell containing one of the cDNA
XX sequences, and a purified antibody capable of binding to one of the
XX secreted proteins. Also contained in the invention are methods for
XX storing the sequence data on a computer system, and a method for
XX identifying features of the cDNA sequences using a computer programme.
XX The cDNAs are useful for expressing secreted proteins or fragments to
XX obtain antibodies capable of specifically binding to the secreted
XX proteins. The cDNAs may also be useful in diagnostic, forensic, gene
XX therapy and chromosome mapping procedures and may be used to design
XX expression vectors and secretion vectors. The proteins of the invention
XX may be used to treat diseases including cancer, autoimmune diseases,
XX cardiovascular disorders, cystic fibrosis, hypothyroidism, immunological
XX disorders, amyloidosis, brain disorders, skeletal muscle disorders, eye
XX disorders, obesity, mitochondrialcytopathies, diabetes, atherosclerosis,
XX neurodegenerative disorders, graft rejection, Alzheimer's disease,
XX dementia, hyperlipidaemia, septic shock and impotence.
XX
XX Sequence 249 AA;
XX
XX Query Match 100.0%; Score 46; DB 21; Length 249;
XX Best Local Similarity 100.0%; Pred. No. 0.91;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 ETDPOTHQ 8
XX |
XX Db 159 etdpqthq 166
XX
XX RESULT 9
XX AAB75351
XX ID AAB75351 standard; protein: 249 AA.
XX
XX AC AAB75351;
XX
XX 05-APR-2001 (first entry)
XX Human secreted protein #10.
XX
XX Secreted protein; prevention; treatment; diagnosis; disease;
XX infection.
XX
XX Homo sapiens.
XX WO200100806-A2.
XX
XX 04-JAN-2001.
XX
XX 21-JUN-2000; 2000WO-IB00951.
XX
XX 25-JUN-1999; 99US-0141032.
XX 21-DEC-1999; 99US-0469099.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Bougueleret L, Jobert S;
XX
XX WPI; 2001-071487/08.
XX
XX 49 Secreted proteins and the nucleic acids encoding them, useful in
XX gene therapy and for detecting similar sequences in samples -
XX
XX Claim 10; Page 281; 307pp; English.
XX
XX The present invention relates to 49 Secreted proteins and the cDNAs
XX encoding them. The protein and nucleic acids may be used in the
XX prevention, treatment and diagnosis of diseases associated with
XX inappropriate protein expression.
XX
XX Sequence 249 AA;
XX
XX Query Match 100.0%; Score 46; DB 22; Length 249;
XX Best Local Similarity 100.0%; Pred. No. 0.91;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 ETDPOTHQ 8
XX |
XX Db 159 etdpqthq 166
XX
XX RESULT 10
XX AAM25745
XX ID AAM25745 standard; Protein: 260 AA.
XX
XX AC AAM25745;
XX
XX 16-OCT-2001 (first entry)
XX Human protein sequence SEQ ID NO:1260.
XX
XX Human; cancer; ulcer; HIV infection; Human immunodeficiency virus;
XX antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
XX antibacterial; endocrine; cardiant; central nervous system; virucide;
XX anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
XX antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;
XX dermatological; antiallergic; antisthmatic; antidiabetic; cytostatic;
XX neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
XX immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
XX antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
XX cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
XX genetic disease; haematopoietic disorder; platelet disorder; asthma;
XX thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
XX allergic rhinitis; diabetes; multiple sclerosis; depression;
XX Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
XX neurological disorder.
XX

```

OS Homo sapiens.
 PN WO200153455-A2.
 XX
 PD 26-JUL-2001.
 XX
 PF 22-DEC-2000; 2000WO-US35017.
 XX
 PR 23-DEC-1999; 99US-0471275.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 XX
 PR 25-APR-2000; 2000US-0552317.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 PI WPI: 2001-457603/49.
 XX
 DR N-PSDB; AAH99686.
 XX
 PT Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
 XX
 PS Claim 20: Page 260; 1217pp; English.
 XX
 CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antihaemic; antiaggregant; haemostatic; vulnery;
 CC antitumor; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.
 XX
 SQ Sequence 260 AA;

Query Match 89.1%; Score 41; DB 22; Length 260;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDPTQH 7
 Db 170 etdptqh 176
 |||||

RESULT 11
 ABG25271
 ID ABG25271 standard; Protein: 88 AA.
 XX
 AC ABG25271;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #25262.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN

PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 PI WPI: 2001-639362/73.
 XX
 DR N-PSDB; AAS89458.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20: SEQ ID No 55630; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of antigens and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG0377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 88 AA;

Query Match 78.3%; Score 36; DB 22; Length 88;
 Best Local Similarity 85.7%; Pred. No. 21;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TDPQTHQ 8
 Db 32 tdpdthq 38
 |||||

RESULT 12
 ABG06840
 ID ABG06840 standard; Protein: 119 AA.
 XX
 AC ABG06840;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #6831.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.

```

XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YF;
XX WPI; 2001-639362/73.
XX DR N-PSDB: AAS71027.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID No 37199; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 119 AA;

Query Match 78.3%; Score 36; DB 22; Length 119;
Best Local Similarity 85.7%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TDPQTHQ 8
DB 94 tdpdthq 100

RESULT 13
ABG04630
ID ABG04630 standard; Protein; 125 AA.
XX AC
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #4621.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD

```

```

PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YF;
XX WPI; 2001-639362/73.
XX DR N-PSDB: AAS68817.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID No 34989; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 125 AA;

Query Match 78.3%; Score 36; DB 22; Length 125;
Best Local Similarity 85.7%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TDPQTHQ 8
DB 6 tdpdthq 12

RESULT 14
ABG23981
ID ABG23981 standard; Protein; 161 AA.
XX AC
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #23972.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.

```

XX 30-MAR-2001; 2001WO-US08631.
 PF XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX XX
 PA (HYSE-) HYSEQ INC.
 XX XX
 PI Drmanac RT, Liu C, Tang YT;
 XX XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS88168.
 XX XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX XX
 PS Claim 20; SEQ ID No 54340; 103pp: English.
 XX XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX XX
 SQ Sequence 161 AA;

Query Match 78.3%; Score 36; DB 22; Length 161;
 Best Local Similarity 85.7%; Pred. No. 40;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TDPOTHQ 8
 ||| |||
 Db 150 tdpthq 156

RESULT 15
 ABG23284
 ID ABG23284 standard; Protein; 181 AA.

XX AC ABG23284;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #23275.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO2001/5067-A2.

XX PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.
 XX XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX XX
 PA (HYSE-) HYSEQ INC.
 XX XX
 PI Drmanac RT, Liu C, Tang YT;
 XX XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS87471.
 XX XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX XX
 PS Claim 20; SEQ ID No 53643; 103pp: English.
 XX XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX XX
 SQ Sequence 181 AA;

Query Match 78.3%; Score 36; DB 22; Length 181;
 Best Local Similarity 85.7%; Pred. No. 45;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TDPOTHQ 8
 ||| |||
 Db 125 tdpthq 131

Search completed: August 7, 2002, 06:45:27
 Job time: 109 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 06:43:38 ; Search time 12.96 Seconds
(without alignments)
15.078 Million cell updates/sec

Title: US-10-020-139-2_COPY_159_166

Perfect score: 46

Sequence: 1 ETPDQTHQ 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	69.6	18	2	US-09-017-205-51
2	32	69.6	18	2	US-09-017-205-52
3	31	67.4	9	3	US-08-159-339A-1190
4	31	67.4	9	3	US-08-159-339A-1192
5	31	67.4	9	3	US-08-159-339A-1193
6	31	67.4	9	3	US-08-159-339A-1194
7	31	67.4	9	3	US-08-159-339A-1200
8	31	67.4	391	5	PCT-US95-15696-2
9	31	67.4	425	2	US-08-951-148-3
10	31	67.4	425	2	US-09-165-234-3
11	31	67.4	425	2	US-09-274-570-3
12	31	67.4	675	3	US-08-947-965-76
13	31	67.4	676	3	US-08-947-965-71
14	31	67.4	724	4	US-09-307-143-6
15	31	67.4	732	2	US-08-533-669A-18
16	31	67.4	732	4	US-09-307-143-4
17	31	67.4	1456	1	US-08-803-973-2
18	31	67.4	1456	1	US-08-803-972-2
19	31	67.4	2257	1	US-08-611-107-10
20	31	67.4	2257	2	US-08-422-560A-10
21	31	67.4	2257	4	US-08-468-793-10
22	30	65.2	209	4	US-08-914-999-14
23	30	65.2	212	4	US-08-914-999-17
24	30	65.2	233	4	US-08-914-999-18
25	30	65.2	238	4	US-08-914-999-15
26	30	65.2	258	4	US-08-914-999-16
27	30	65.2	380	2	US-08-773-870-5

28 30 65.2 540 3 US-08-964-268-6 Sequence 6, Appl
29 30 65.2 724 4 US-08-914-999-4 Sequence 4, Appl
30 30 65.2 725 4 US-08-914-999-2 Sequence 2, Appl
31 30 65.2 732 4 US-08-914-999-8 Sequence 8, Appl
32 30 65.2 760 4 US-08-914-999-12 Sequence 12, Appl
33 30 65.2 768 4 US-08-914-999-10 Sequence 10, Appl
34 30 65.2 803 4 US-09-063-035-2 Sequence 2, Appl
35 30 65.2 1146 4 US-08-914-999-6 Sequence 6, Appl
36 30 65.2 1346 2 US-08-635-121-2 Sequence 2, Appl
37 29 63.0 9 3 US-08-159-339A-1189 Sequence 1189, Ap
38 29 63.0 221 4 US-08-896-933-29 Sequence 29, Appl
39 29 63.0 221 4 US-09-314-235-29 Sequence 2, Appl
40 29 63.0 292 2 US-08-879-260-2 Sequence 14, Appl
41 29 63.0 316 4 US-09-413-814-14 Sequence 56, Appl
42 29 63.0 365 4 US-09-004-838-133 Sequence 16, Appl
43 29 63.0 388 2 US-08-282-197C-56 Sequence 23, Appl
44 29 63.0 402 4 US-09-347-801-16
45 29 63.0 407 4 US-09-347-801-23

ALIGNMENTS

RESULT 1
US-09-017-205-51
; Sequence 51, Application US/09017205
; Patent No. 5965357
; GENERAL INFORMATION:
; APPLICANT: Marsden, Howard S
; TITLE OF INVENTION: PEPTIDE STRUCTURES AND THEIR USE IN
; TITLE OF INVENTION: DIAGNOSIS OF HERPES SIMPLEX VIRUS TYPE 2
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 8th Floor, 1100 No. 5965357th Glebe Road
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/017,205
; FILING DATE: 02-FEB-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29,009
; REFERENCE/DOCKET NUMBER: 604-436
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)816-4000
; TELEFAX: (703)816-4100
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide from HSV-2 glycoprotein G
; FRAGMENT TYPE: internal
US-09-017-205-51

Query Match 69.6%; Score 32; DB 2; Length 18;
Best Local Similarity 83.3%; Pred. No. 6.2;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TDPQTH 7

Db 13 TDPKTH 18

```

RESULT 2
US-09-017-205-52
: Sequence 52, Application US/09017205
: Patent No. 5965357
: GENERAL INVENTOR: Marsden, Howard S
: APPLICANT: Marsden, Howard S
: TITLE OF INVENTION: PEPTIDE STRUCTURES AND THEIR USE IN
: TITLE OF INVENTION: DIAGNOSIS OF HERPES SIMPLEX VIRUS TYPE 2
: NUMBER OF SEQUENCES: 86
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nixon & Vanderhye PC
: STREET: 8th Floor, 1100 No. 5965357th Glebe Road
: CITY: Arlington
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22201-4714
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/017,205
: FILING DATE: 02-FEB-1998
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Mitchard, Leonard C
: REGISTRATION NUMBER: 29,009
: REFERENCE/DOCKET NUMBER: 504-436
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)816-4000
: TELEFAX: (703)816-4100
: INFORMATION FOR SEQ ID NO: 52:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 18 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide from HSV-2 glycoprotein G
: FRAGMENT TYPE: internal
: US-09-017-205-52

Query Match 59.6%; Score 32; DB 2; Length 18;
Best Local Similarity 53.3%; Pred. No. 6.2;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TDPQTH 7
Db 3 TDPKTH 8

RESULT 3
US-08-159-339A-1190
: Sequence 1190, Application US/08159339A
: Patent No. 6037135
: GENERAL INFORMATION:
: APPLICANT: Kubo, Ralph T.
: APPLICANT: Grey, Howard M.
: APPLICANT: Sette, Alessandro
: APPLICANT: Celis, Esteban
: TITLE OF INVENTION: HLA Binding peptides and Their
: TITLE OF INVENTION: Uses
: NUMBER OF SEQUENCES: 1254
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette

```

; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 018623-005030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX:
; INFORMATION FOR SEQ ID NO: 1192:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-159-339A-1192

Query Match 67.4%; Score 31; DB 3; Length 9;
Best Local Similarity 71.4%; Pred. No. 1.7e+05;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDPOTH 7
Db 1 ETDPTSH 7

RESULT 5
US-08-159-339A-1193
; Sequence 1193, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 018623-005030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX:

; INFORMATION FOR SEQ ID NO: 1193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-159-339A-1193

Query Match 67.4%; Score 31; DB 3; Length 9;
Best Local Similarity 71.4%; Pred. No. 1.7e+05;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDPOTH 7
Db 1 ETDPTSH 7

RESULT 6
US-08-159-339A-1194
; Sequence 1194, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 018623-005030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX:
; INFORMATION FOR SEQ ID NO: 1194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-159-339A-1194

Query Match 67.4%; Score 31; DB 3; Length 9;

Best Local Similarity 71.4%; Pred. No. 1.7e+05;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDPTH 7
111111
Db 1 ETDPTH 7

RESULT 7
US-08-159-339A-1200
; Sequence 1200, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 018623-0050300S
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX:
; INFORMATION FOR SEQ ID NO: 1200:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-159-339A-1200

Query Match 67.4%; Score 31; DB 3; Length 9;
Best Local Similarity 71.4%; Pred. No. 1.7e+05;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDPTH 7
111111
Db 1 ETDPTH 7

RESULT 8
PCT-US95-15696-2
; Sequence 2, Application PC/TUS9515696

; GENERAL INFORMATION:
; APPLICANT: President and Fellows of Harvard College
; TITLE OF INVENTION: SINGLE CHAIN T-CELL RECEPTOR
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: Maine
; COUNTRY: USA
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/15696
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/349,915
; FILING DATE: 06-DEC-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: HU-9404 WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 207-363-0558
; TELEFAX: 207-363-0528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-15696-2

Query Match 67.4%; Score 31; DB 5; Length 391;
Best Local Similarity 71.4%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TDPQTH 8
111111
Db 308 TDPQTH 314

RESULT 9
US-08-951-148-3
; Sequence 3, Application US/08951148
; Patent No. 5871973
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: CELL DIVISION REGULATORS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:


```

; APPLICATION NUMBER: US/08/951,148
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0407 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 425 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LVZNNOT01
; CLONE: 348429
; US-08-951-148-3

```

Query Match 67.4%; Score 31; DB 2; Length 425;
Best Local Similarity 71.4%; Pred. No. 2.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 ETDPOTH 7
    1:11111
Db 67 ESDPATH 73

```

```

RESULT 10
US-09-165-234-3
; Sequence 3, Application US/09165234
; Patent No. 5528899
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: CELL DIVISION REGULATORS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/165,234
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/951,148
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0407 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:

```

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 425 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LVZNNOT01
; CLONE: 348429
; US-09-165-234-3

```

Query Match 67.4%; Score 31; DB 2; Length 425;
Best Local Similarity 71.4%; Pred. No. 2.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 ETDPOTH 7
    1:11111
Db 67 ESDPATH 73

```

```

RESULT 11
US-09-274-570-3
; Sequence 3, Application US/09274570
; Patent No. 6121019
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: CELL DIVISION REGULATORS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/274,570
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/951,148
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0407 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 425 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LVZNNOT01
; CLONE: 348429
; US-09-274-570-3

```

Query Match 67.4%; Score 31; DB 3; Length 425;
Best Local Similarity 71.4%; Pred. No. 2.5e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ETDPOTH 7
Db 67 ESDPATH 73

RESULT 12
US-08-947-965-76
; Sequence 76, Application US/08947965A
; Patent No. 6004790
; GENERAL INFORMATION:
; APPLICANT: Dijkhuizen, Lubbert
; APPLICANT: Dijkstra, Bauke
; APPLICANT: Andersen, Carsten
; TITLE OF INVENTION: Cyclomaltodextrin Glucanotransferase
; FILE REFERENCE: 4285.204-US
; CURRENT APPLICATION NUMBER: US/08/947,965A
; EARLIER FILING DATE: 1997-10-09
; EARLIER FILING DATE: 1995-04-21
; EARLIER FILING DATE: 1995-04-21
; EARLIER FILING DATE: 1995-10-17
; EARLIER FILING DATE: 1995-11-16
; EARLIER FILING DATE: 1996-04-22
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 76
; LENGTH: 675
; TYPE: PRT
; ORGANISM: Bacillus ohbensis
US-08-947-965-76

Query Match 67.4%; Score 31; DB 3; Length 675;
Best Local Similarity 71.4%; Pred. No. 4e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ETDPOTH 7
Db 257 EVDPONH 263

RESULT 13
US-08-947-965-71
; Sequence 71, Application US/08947965A
; Patent No. 6004790
; GENERAL INFORMATION:
; APPLICANT: Dijkhuizen, Lubbert
; APPLICANT: Dijkstra, Bauke
; APPLICANT: Andersen, Carsten
; APPLICANT: Osten, Claus von der
; TITLE OF INVENTION: Cyclomaltodextrin Glucanotransferase
; FILE REFERENCE: 4285.204-US
; CURRENT APPLICATION NUMBER: US/08/947,965A
; EARLIER FILING DATE: 1997-10-09
; EARLIER FILING DATE: 1995-04-21
; EARLIER FILING DATE: 1995-10-17
; EARLIER FILING DATE: 1995-11-16
; EARLIER FILING DATE: 1996-04-22
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 71
; LENGTH: 676

; TYPE: PRT
; ORGANISM: Bacillus sp.
US-08-947-965-71

Query Match 67.4%; Score 31; DB 3; Length 676;
Best Local Similarity 71.4%; Pred. No. 4.1e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ETDPOTH 7
Db 259 EVDPONH 265

RESULT 14
US-09-307-143-6
; Sequence 6, Application US/093071143
; Patent No. 6335157
; GENERAL INFORMATION:
; APPLICANT: Gonzalez C.
; APPLICANT: Lange, B.
; TITLE OF INVENTION: METHODS BASED ON LOCALIZATION OF HSP90 TO THE
; FILE REFERENCE: 9882-003
; CURRENT APPLICATION NUMBER: US/09/307,143
; CURRENT FILING DATE: 1999-05-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-307-143-6

Query Match 67.4%; Score 31; DB 4; Length 724;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DPQTH 7
Db 672 DPQTH 676

RESULT 15
US-08-533-669A-18
; Sequence 18, Application US/08533669A
; Patent No. 5834592
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,669A
; FILING DATE: 22-SEP-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.420

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 732 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-533-669A-18

Query Match 67.4%; Score 31; DB 2; Length 732;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DPQTH 7
|
|
|
|
Db 680 DPQTH 684

Search completed: August 7, 2002, 06:44:51
Job time: 73 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: August 6, 2002, 17:05:16 ; Search time 66.51 seconds
(without alignments)
13,003 Million cell updates/sec

Title: US-10-020-139-2_COPY_237_245

Perfect score: 51

Sequence: 1 DNFQHKTQL 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	72.5	498	1	VHIV34
2	37	72.5	498	1	VHIVXL
3	37	72.5	498	1	VHIV68
4	37	72.5	498	1	VHIV61
5	37	72.5	498	1	VHIVAK
6	37	72.5	498	1	VHIVN5
7	37	72.5	498	1	VHIVN8
8	37	72.5	498	1	VHIVN7
9	37	72.5	498	1	VHIVX2
10	37	72.5	498	1	VHIVX3
11	37	72.5	498	1	VHIVX4
12	37	72.5	498	1	VHIVX5
13	37	72.5	498	1	VHIVX6
14	37	72.5	498	1	VHIVN9
15	37	72.5	498	1	VHIVN6
16	37	72.5	498	1	VHIVN3
17	37	72.5	498	1	VHIVN2
18	37	72.5	498	1	VHIVN1
19	37	72.5	498	1	VHIVC1
20	37	72.5	498	1	VHIVM1
21	37	72.5	498	1	A60028
22	37	72.5	498	1	VHIVA3
23	37	72.5	498	1	VHIVA6
24	37	72.5	498	1	VHIVX1
25	37	72.5	498	1	VHIVN8
26	37	72.5	498	1	VHIVN4
27	37	72.5	498	1	B36754
28	37	72.5	498	2	A42757
29	37	72.5	498	2	A42757

30 37 72.5 498 2 S34418 nucleoprotein - in
31 36 70.6 104 2 H65096 hypothetical 12.1
32 36 70.6 104 2 E91124 hypothetical prote
33 36 70.6 104 2 D85969 hypothetical prote
34 36 70.6 482 2 D97162 UDP-N-acetylmuram
35 36 70.6 1246 2 AC2372 hypothetical prote
36 35 68.6 271 2 AG0315 conserved hypothet
37 35 68.6 577 2 D91339 membrane protein
38 35 68.6 577 2 F65202 hypothetical 66.6
39 35 68.6 577 2 A86087 hypothetical prote
40 35 68.6 577 2 AE0936 probable membrane
41 35 68.6 660 2 C90343 conserved hypothet
42 35 68.6 768 2 T45876 hypothetical prote
43 34 66.7 219 2 H90865 probable beta-phos
44 34 66.7 219 2 A85753 probable beta-phos
45 34 66.7 231 2 S73706 hypothetical prote

ALIGNMENTS

RESULT 1

VHIV34

nucleoprotein - Influenza A virus (strain A/PR/8/34)

C:Species: Influenza A virus

C>Date: 18-Aug-1982 #sequence_revision 17-Dec-1982 #text_change 24-Feb-1995

C:Accession: A94327; A94590; A04077

R:Winter, G.; Fields, S.

Virology 114, 423-428, 1981

A:Title: The structure of the gene encoding the nucleoprotein of human influenza vir

A:Reference number: A94327; MUID:82041445

A:Accession: A94327

A:Molecule type: genomic RNA

A:Residues: 1-498 <WIN>

R:Van Rompay, L.; Min Jou, W.; Huylebroeck, D.; Devos, R.; Fiers, W.

Submitted to the Atlas, June 1982

A:Reference number: A94590

A:Accession: A94590

A:Molecule type: genomic RNA

A:Residues: 1-246, 'N', 248-352, 'L', 354-424, 'I', 426-429, 'N', 431-498 <VAN>

A:Experimental source: substrain HON1

C:Genetics:

A:Map position: segment 5

C:Superfamily: Influenza virus nucleoprotein

Query Match 72.5% Score 37; DB 1: Length 498;
Best local Similarity 66.7% Pred. No. 19;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNFQHKTQL 9

DB 320 ENPAHKSQL 328

RESULT 2

VHIVXL

nucleoprotein - Influenza A virus (strain X/Leningrad/54/1 [H1N1])

C:Species: Influenza A virus

A:Note: host Homo sapiens (man)

C>Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 16-Jul-1999

C:Accession: JN0394

R:Beklemishev, A.B.; Blinov, V.M.; Vassilenko, S.K.; Golovin, S.Y.; Karginov, V.A.;

Bioorg. Khim. 11, 636-640, 1985

A:Title: Synthesis, cloning and sequencing of a full-length DNA copy of NP gene of t

A:Reference number: JN0394; MUID:85307106

A:Accession: JN0394

A:Molecule type: genomic RNA

A:Residues: 1-498 <BEM>

A:Cross-references: GB:M38279; NID:g324691; PIDN:AAA43459.1; PID:g324692

A:Note: the authors translated the codon GUG for residue 67 as Ala

C:Genetics:

A:Gene: NP

A:Map position: segment 5
 C:Superfamily: influenza virus nucleoprotein
 C:Keywords: nucleoprotein; phosphate (Ser)
 F:176,345/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 72.5%; Score 37; DB 1; Length 498;
 Best Local Similarity 66.7%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTL 9
 :||| |||

Db 320 ENPAHKSQ 328

RESULT 3

VHIV68

nucleoprotein - influenza A virus (strain A/NT/60/68)

C:Species: influenza A virus

C>Date: 17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change 16-Jul-1999

R:Accession: A04078; F36754

R:Huddleston, J.A.; Brownlee, G.G.

Nucleic Acids Res. 10, 1029-1038, 1982

A:Title: The sequence of the nucleoprotein gene of human influenza A virus, strain A/NT/60/68

A:Reference number: A04078; MUID:82150233

A:Molecule type: genomic RNA

A:Residues: 1-498 <NUD>

A:Cross-references: GB:J02137; MID:g324703; PIDN:AAA43465.1; PID:g324704

A:Experimental source: strain A/NT/60/68

R:Altmueller, A.; Fitch, W.M.; Scholtissek, C.

J. Gen. Virol. 70, 2111-2119, 1989

A:Title: Biological and genetic evolution of the nucleoprotein gene of human influenza A

A:Reference number: A36754; MUID:89361370

A:Accession: F36754

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-498 <ALT>

A:Experimental source: strain A/NT/60/68 [H3N2]

C:Genetics:

A:Gene: NP

A:Map position: segment 5

C:Superfamily: influenza virus nucleoprotein

Query Match 72.5%; Score 37; DB 1; Length 498;
 Best Local Similarity 66.7%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTL 9
 :||| |||

Db 320 ENPAHKSQ 328

RESULT 4

VHIV61

nucleoprotein - influenza A virus (strain A/Ann Arbor/6/60 [H2N2])

N:Alternate names: NP protein

C:Species: influenza A virus

C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 08-Apr-1994

R:Accession: D31831

R:Cox, N.J.; Kitame, F.; Kendal, A.P.; Maassab, H.F.; Naeve, C.

Virol. 167, 554-567, 1988

A:Title: Identification of sequence changes in the cold-adapted, live attenuated influen

A:Reference number: A31831; MUID:89073759

A:Accession: D31831

A:Molecule type: genomic RNA

A:Residues: 1-498 <COX>

C:Genetics:

A:Map position: segment 5

C:Superfamily: influenza virus nucleoprotein

C:Keywords: nucleoprotein

Query Match 72.5%; Score 37; DB 1; Length 498;
 Best Local Similarity 66.7%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTL 9
 :||| |||

Db 320 ENPAHKSQ 328

RESULT 5

VHIV7

nucleoprotein - influenza A virus (strain A/Udorn/307/72)

C:Species: influenza A virus

A:Note: host Homo sapiens (man)

C>Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jul-1999

C:Accession: B25612; G36754

R:Buckler-White, A.J.; Murphy, B.R.

Virol. 155, 345-355, 1986

A:Title: Nucleotide sequence analysis of the nucleoprotein gene of an avian and a hum

A:Reference number: A94345; MUID:87071656

A:Accession: B25612

A:Molecule type: mRNA

A:Residues: 1-498 <BUC>

A:Cross-references: GB:M14922; MID:g325098; PIDN:AAA43686.1; PID:g325099

A:Experimental source: strain A/Udorn/307/72

R:Altmueller, A.; Fitch, W.M.; Scholtissek, C.

J. Gen. Virol. 70, 2111-2119, 1989

A:Title: Biological and genetic evolution of the nucleoprotein gene of human influen

A:Reference number: A36754; MUID:89361370

A:Accession: G36754

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-498 <ALT>

A:Experimental source: strain A/Udorn/307/72 [H3N2]

C:Genetics:

A:Gene: NP

A:Map position: segment 5

C:Superfamily: influenza virus nucleoprotein

C:Keywords: nucleoprotein

Query Match 72.5%; Score 37; DB 1; Length 498;
 Best Local Similarity 66.7%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTL 9
 :||| |||

Db 320 ENPAHKSQ 328

RESULT 6

VHIVAK

nucleoprotein - influenza A virus (strain A/Kiev/59/79 [H1N1])

C:Species: influenza A virus

A:Note: host Homo sapiens (man)

C>Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 16-Jul-1999

C:Accession: JN0399; S09650

R:Beklemishev, A.B.; Blynov, V.M.; Vassilenko, S.K.; Golovin, S.Y.; Karginov, V.A.; M

Biorg. Khim. 12, 369-374, 1986

A:Title: Nucleotide sequence of a full-length DNA copy of the influenza virus A/Kiev/

A:Reference number: S09650; MUID:86186950

A:Accession: JN0399

A:Molecule type: genomic RNA

A:Residues: 1-498 <BBK>

A:Cross-references: EMBL:X51972; MID:g60820; PIDN:CAA36234.1; PID:g60821

C:Genetics:

A:Gene: NP

A:Map position: segment 5

C:Superfamily: influenza virus nucleoprotein

C:Keywords: nucleoprotein; phosphate (Ser)

F:176,345/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 72.5%; Score 37; DB 1; Length 498;
 Best Local Similarity 66.7%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPOHKTOL 9
 :||| ||:||
 Db 320 ENPAHKSOL 328

RESULT 7
 VIHIVS
 nucleoprotein - influenza A virus (strain A/Hong Kong/5/83 [H3N2])
 N:Alternate names: NP protein
 C:Species: influenza A virus
 C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
 C:Accession: E31470; H36755
 R:Gammelinn, M.; Mandler, J.; Scholtissek, C.
 Virol. 170, 71-80, 1989
 A:Title: Two subtypes of nucleoproteins (NP) of influenza A viruses.
 A:Reference number: A31470; MUID:89243210
 A:Accession: E31470
 A:Molecule type: genomic RNA
 A:Residues: 1-498 <GAM>
 A:Cross-references: GB:M22577; GB:J04339; NID:g324245; PIDN:AAA43241.1; PID:g324246
 R:Altmueller, A.; Fitch, W.M.; Scholtissek, C.
 J. Gen. Virol. 70, 2111-2119, 1989
 A:Title: Biological and genetic evolution of the nucleoprotein gene of human influenza A
 A:Reference number: A36754; MUID:89361370
 A:Accession: B36755
 A>Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1-498 <ALT>
 C:Genetics:
 A:Gene: NP
 A:Map position: segment 5
 C:Superfamily: influenza virus nucleoprotein
 C:Keywords: nucleoprotein

Query Match 72.5%; Score 37; DB 1; Length 498;
 Best Local Similarity 66.7%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPOHKTOL 9
 :||| ||:||
 Db 320 ENPAHKSOL 328

RESULT 8
 VIHIVH
 nucleoprotein - influenza A virus (strain A/Hong Kong/1/68 [H3N2])
 C:Species: influenza A virus
 C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999
 C:Accession: S05502
 R:Reinhardt, U.; Mandler, J.; Scholtissek, C.
 Nucleic Acids Res. 17, 6721, 1989
 A:Title: Sequence of the nucleoprotein (NP) gene of the influenza A virus reassortant 81
 A:Reference number: S05502; MUID:89385995
 A:Accession: S05502
 A:Molecule type: DNA
 A:Residues: 1-498 <REI>
 A:Cross-references: EMBL:X15890; NID:g60477; PIDN:CAA33899.1; PID:g60478
 C:Genetics:
 A:Gene: NP
 A:Map position: segment 5
 C:Superfamily: influenza virus nucleoprotein
 C:Keywords: nucleoprotein

Query Match 72.5%; Score 37; DB 1; Length 498;
 Best Local Similarity 66.7%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPOHKTOL 9
 :||| ||:||
 Db 320 ENPAHKSOL 328

RESULT 8
 VIHIVH
 nucleoprotein - influenza A virus (strain A/Hong Kong/1/68 [H3N2])
 C:Species: influenza A virus
 C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999
 C:Accession: S05502
 R:Reinhardt, U.; Mandler, J.; Scholtissek, C.
 Nucleic Acids Res. 17, 6721, 1989
 A:Title: Sequence of the nucleoprotein (NP) gene of the influenza A virus reassortant 81
 A:Reference number: S05502; MUID:89385995
 A:Accession: S05502
 A:Molecule type: DNA
 A:Residues: 1-498 <REI>
 A:Cross-references: EMBL:X15890; NID:g60477; PIDN:CAA33899.1; PID:g60478
 C:Genetics:
 A:Gene: NP
 A:Map position: segment 5
 C:Superfamily: influenza virus nucleoprotein
 C:Keywords: nucleoprotein

Query Match 72.5%; Score 37; DB 1; Length 498;
 Best Local Similarity 66.7%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPOHKTOL 9
 :||| ||:||
 Db 320 ENPAHKSOL 328

RESULT 11
 VIHIVX3
 nucleoprotein - influenza A virus (strain A/USSR/90/77 [H1N1])
 N:Alternate names: NP protein

QY 1 DNPOHKTOL 9
 :||| ||:||
 Db 320 ENPAHKSOL 328

RESULT 9
 VIHIV7
 nucleoprotein - influenza A virus (strain A/swine/Hong Kong/6/76 [H3N2])
 N:Alternate names: NP protein
 C:Species: influenza A virus
 C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
 C:Accession: G31470
 R:Gammelinn, M.; Mandler, J.; Scholtissek, C.
 Virol. 170, 71-80, 1989
 A:Title: Two subtypes of nucleoproteins (NP) of influenza A viruses.
 A:Reference number: A31470; MUID:89243210
 A:Accession: G31470
 A:Molecule type: genomic RNA
 A:Residues: 1-498 <GAM>
 A:Cross-references: GB:M22571; GB:J04339; NID:g325058; PIDN:AAA43668.1; PID:g325059
 C:Genetics:
 A:Gene: NP
 A:Map position: segment 5
 C:Superfamily: influenza virus nucleoprotein
 C:Keywords: nucleoprotein

Query Match 72.5%; Score 37; DB 1; Length 498;
 Best Local Similarity 66.7%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPOHKTOL 9
 :||| ||:||
 Db 320 ENPAHKSOL 328

RESULT 10
 VIHIVX2
 nucleoprotein - influenza A virus (strain A/Fort Warren/1/50 [H1N1])
 N:Alternate names: NP protein
 C:Species: influenza A virus
 A:Note: host Homo sapiens (man)
 C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jun-2000
 C:Accession: C36754
 R:Altmueller, A.; Fitch, W.M.; Scholtissek, C.
 J. Gen. Virol. 70, 2111-2119, 1989
 A:Title: Biological and genetic evolution of the nucleoprotein gene of human influer
 A:Reference number: A36754; MUID:89361370
 A:Accession: C36754
 A:Molecule type: mRNA
 A:Residues: 1-498 <ALT>
 A:Cross-references: GB:D00601; NID:g221287; PIDN:BAA00477.1; PID:g221288
 C:Genetics:
 A:Gene: NP
 A:Map position: segment 5
 C:Superfamily: influenza virus nucleoprotein
 C:Keywords: nucleoprotein

Query Match 72.5%; Score 37; DB 1; Length 498;
 Best Local Similarity 66.7%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPOHKTOL 9
 :||| ||:||
 Db 320 ENPAHKSOL 328

RESULT 11
 VIHIVX3
 nucleoprotein - influenza A virus (strain A/USSR/90/77 [H1N1])
 N:Alternate names: NP protein

C:Species: Influenza A virus
A:Note: host Homo sapiens (man)
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 08-Apr-1994
C:Accession: D36754
R:Altmueller, A.; Fitch, W.M.; Scholtissek, C.
J. Gen. Virol. 70, 2111-2119, 1989
A:Title: Biological and genetic evolution of the nucleoprotein gene of human influenza A
A:Reference number: A36754; MUID:89361370
A:Accession: D36754
A:Molecule type: mRNA
A:Residues: 1-498 <ALT>
C:Genetics:
A:Gene: NP
A:Map position: segment 5
C:Superfamily: influenza virus nucleoprotein
C:Keywords: nucleoprotein

Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTLQ 9
DB 320 ENPAHKSQ 328

RESULT 12

VHIVX4

nucleoprotein - influenza A virus (strain A/Brazil/11/78 [H1N1])

N:Alternate names: NP protein

C:Species: Influenza A virus

A:Note: host Homo sapiens (man)

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 08-Apr-1994

C:Accession: E36754

R:Altmueller, A.; Fitch, W.M.; Scholtissek, C.

J. Gen. Virol. 70, 2111-2119, 1989

A:Title: Biological and genetic evolution of the nucleoprotein gene of human influenza A

A:Reference number: A36754; MUID:89361370

A:Accession: E36754

A:Molecule type: mRNA

A:Residues: 1-498 <ALT>

C:Genetics:

A:Gene: NP

A:Map position: segment 5

C:Superfamily: influenza virus nucleoprotein

C:Keywords: nucleoprotein

Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTLQ 9
DB 320 ENPAHKSQ 328

RESULT 13

VHIVX5

nucleoprotein - influenza A virus (strain A/Texas [H3N2])

N:Alternate names: NP protein

C:Species: Influenza A virus

A:Note: host Homo sapiens (man)

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jun-2000

C:Accession: H36754; P00411; P00423

R:Altmueller, A.; Fitch, W.M.; Scholtissek, C.

J. Gen. Virol. 70, 2111-2119, 1989

A:Title: Biological and genetic evolution of the nucleoprotein gene of human influenza A

A:Reference number: A36754; MUID:89361370

A:Accession: H36754

A:Molecule type: mRNA

A:Residues: 1-498 <ALT>

A:Cross-references: GB:D00602; NID:q221289; PIDN:BAA00478.1; PID:q221290
A:Experimental source: strain A/Texas [H3N2]
K:Li, X.S.; Zhao, C.Y.; Gao, H.M.; Zhang, Y.Q.; Ishida, M.; Kanegae, Y.; Endo, A.; Ne
J. Gen. Virol. 73, 1329-1337, 1992
A:Title: Origin and evolutionary characteristics of antigenic reassortant influenza A
A:Reference number: P00408; MUID:92300326
A:Accession: P00411
A:Molecule type: genomic RNA
A:Residues: 23-55 <LIAI>
A:Experimental source: strain A/Hubei/24/89 [H1N2]
A:Accession: P00423
A:Molecule type: genomic RNA
A:Residues: 23-55 <LIA2>
A:Experimental source: strain A/Guizhou/54/89 [H3N2]
C:Genetics:
A:Gene: NP
A:Map position: segment 5
C:Superfamily: influenza virus nucleoprotein
C:Keywords: nucleoprotein

Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTLQ 9
DB 320 ENPAHKSQ 328

RESULT 14

VHIVX6

nucleoprotein - influenza A virus (strain A/California [H1N1])

N:Alternate names: NP protein

C:Species: Influenza A virus

A:Note: host Homo sapiens (man)

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jun-2000

C:Accession: A36755

R:Altmueller, A.; Fitch, W.M.; Scholtissek, C.

J. Gen. Virol. 70, 2111-2119, 1989

A:Title: Biological and genetic evolution of the nucleoprotein gene of human influenza

A:Reference number: A36754; MUID:89361370

A:Accession: A36755

A:Molecule type: mRNA

A:Residues: 1-498 <ALT>

A:Cross-references: GB:D00600; NID:q221285; PIDN:BAA00476.1; PID:q221286

C:Genetics:

A:Gene: NP

A:Map position: segment 5

C:Superfamily: influenza virus nucleoprotein

C:Keywords: nucleoprotein

Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTLQ 9
DB 320 ENPAHKSQ 328

RESULT 15

VHIVX9

nucleoprotein - influenza A virus (strain A/swine/Iowa/1976/31 [H1N1])

N:Alternate names: NP protein

C:Species: Influenza A virus

C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999

C:Accession: I31470

R:Gamelin, M.; Mandler, J.; Scholtissek, C.

Virology 170, 71-80, 1989

A:Title: Two subtypes of nucleoproteins (NP) of influenza A viruses.

A:Reference number: A31470; MUID:89243210

A:Accession: I31470
A:Molecule type: genomic RNA
A:Residues: 1-498 <GAM>
A:Cross-references: GB:M22578; GB:J04339; NID:g325071; PIDN:AAA3676.1; PID:g325072
C:Genetics:
A:Gene: NP
A:Map position: segment 5
C:Superfamily: influenza virus nucleoprotein
C:Keywords: nucleoprotein

Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPOHKTOL 9
DB 320 ENPAKSQL 328

Search completed: August 6, 2002, 17:05:17
Job time: 919 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 17:07:36 ; Search time 32.88 Seconds
(without alignments)
10.598 Million cell updates/sec

Title: US-10-020-139-2_COPY_237_245

Perfect score: 51
Sequence: 1 DNPQHKTL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	72.5	498	1 VNUC_IAANA	P18277 influenza a
2	37	72.5	498	1 VNUC_IAANN	P21433 influenza a
3	37	72.5	498	1 VNUC_IAB37	Q07539 influenza a
4	37	72.5	498	1 VNUC_IAB39	Q08028 influenza a
5	37	72.5	498	1 VNUC_IABRA	P18069 influenza a
6	37	72.5	498	1 VNUC_IABUD	P15660 influenza a
7	37	72.5	498	1 VNUC_IACAL	P18070 influenza a
8	37	72.5	498	1 VNUC_IACKG	P16984 influenza a
9	37	72.5	498	1 VNUC_IACKP	P15661 influenza a
10	37	72.5	498	1 VNUC_IADAU	P26085 influenza a
11	37	72.5	498	1 VNUC_IADBE	P26084 influenza a
12	37	72.5	498	1 VNUC_IADCE	P15662 influenza a
13	37	72.5	498	1 VNUC_IADE1	P26062 influenza a
14	37	72.5	498	1 VNUC_IADE2	P26063 influenza a
15	37	72.5	498	1 VNUC_IADHK	P16978 influenza a
16	37	72.5	498	1 VNUC_IADM2	P15680 influenza a
17	37	72.5	498	1 VNUC_IADMA	P26055 influenza a
18	37	72.5	498	1 VNUC_IADNZ	P15664 influenza a
19	37	72.5	498	1 VNUC_IADU2	P15663 influenza a
20	37	72.5	498	1 VNUC_IADN5	P26071 influenza a
21	37	72.5	498	1 VNUC_IADN5	P18071 influenza a
22	37	72.5	498	1 VNUC_IADN5	P26061 influenza a
23	37	72.5	498	1 VNUC_IADPD	P26061 influenza a
24	37	72.5	498	1 VNUC_IADPR	P12604 influenza a
25	37	72.5	498	1 VNUC_IADG7	Q09159 influenza a
26	37	72.5	498	1 VNUC_IAGU1	P26059 influenza a
27	37	72.5	498	1 VNUC_IAGU2	P15665 influenza a
28	37	72.5	498	1 VNUC_IAGU3	P15666 influenza a
29	37	72.5	498	1 VNUC_IAGU4	P15667 influenza a
30	37	72.5	498	1 VNUC_IAGUA	P15668 influenza a
31	37	72.5	498	1 VNUC_IAGUA	P15669 influenza a
32	37	72.5	498	1 VNUC_IAGUN	P15670 influenza a
33	37	72.5	498	1 VNUC_IAHIC	P26069 influenza a

34	37	72.5	498	1 VNUC_IAHJI	P26068 influenza a
35	37	72.5	498	1 VNUC_IAHLO	P15673 influenza a
36	37	72.5	498	1 VNUC_IAHMI	P16979 influenza a
37	37	72.5	498	1 VNUC_IAH01	P22435 influenza a
38	37	72.5	498	1 VNUC_IAH02	P16982 influenza a
39	37	72.5	498	1 VNUC_IAHPR	P16980 influenza a
40	37	72.5	498	1 VNUC_IAHTE	P15674 influenza a
41	37	72.5	498	1 VNUC_IAKIE	P16314 influenza a
42	37	72.5	498	1 VNUC_IAKIT	O91743 influenza a
43	37	72.5	498	1 VNUC_IALEN	P31609 influenza a
44	37	72.5	498	1 VNUC_IAMAA	P15675 influenza a
45	37	72.5	498	1 VNUC_IAMAN	P06826 influenza a

ALIGNMENTS

RESULT	1
VNUC_IAANA	
ID	VNUC_IAANA
AC	P18277; STANDARD: PRT: 498 AA.
DT	01-NOV-1990 (Rel. 16, Created)
DT	01-NOV-1990 (Rel. 16, Last sequence update)
DT	01-JUN-1994 (Rel. 29, Last annotation update)
DE	Nucleoprotein.
GN	Np.
OS	Influenza A virus (strain A/Anas acuta/Primorje/695/76).
OC	Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;
OC	Influenza virus A and B group; Influenza A viruses;
OC	Influenza A virus.
OX	NCRI_TaxID=11323;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=91288235; PubMed=2062661;
RA	Mandler J., Kureli M.S., Ludwig S., Herget M.E., Scholtissek C.;
RT	Sequence of the nucleoprotein (NP) gene of the Influenza A virus
RT	acuta/Primorje/695/76 (H2N3) virus."
RL	Nucleic Acids Res. 19:3456-3456(1991).
CC	!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: M36812; AAA43129.1; .
DR	PIR: S34418; S34418.
DR	InterPro: IPR002141; Flu_NP.
DR	Pfam: PF00506; Flu_NP; 1.
KW	Nucleoprotein.
SQ	SEQUENCE 498 AA: 56348 MW: 809E162F61077244 CRC64;

Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY	1 DNPQHKTL 9
	:
Db	320 ENPAKSQL 328

RESULT	2
VNUC_IAANN	
ID	VNUC_IAANN
AC	P21433; P80881; STANDARD: PRT: 498 AA.
DT	01-MAY-1991 (Rel. 18, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	01-NOV-1997 (Rel. 35, Last annotation update)
DE	Nucleoprotein.

```

GN NP.
OS Influenza A virus (strain A/Ann Arbor/6/60).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=135322;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89073759; PubMed-2974219;
RA Cox N.J., Kitame F., Kendal A.P., Maassab H.F., Naeve C.;
RT "Identification of sequence changes in the cold-adapted, live
RL attenuated influenza vaccine strain, A/Ann Arbor/6/60 (H2N2).";
RL Virology 167:554-567(1988).
CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M23976; AAA43451.1; -
DR FIRM; D31831; VHI61.
DR InterPro: IPR002141; Flu_NP.
DR Pfam: PF00506; Flu_NP; 1.
KW Nucleoprotein.
FT VARIANT 23 23 N -> T (IN COLD-ADAPTED ISOLATE).
FT VARIANT 34 34 G -> D (IN COLD-ADAPTED ISOLATE).
FT SEQUENCE 498 AA: 55963 MW: E51D08494E9FA075 CRC64;
SQ
Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 DNPQHKTLQ 9
Db 320 ENPAHKSQ 328
RESULT 3
VNUC_IAB37
ID VNUC_IAB37 STANDARD; PRT; 498 AA.
AC Q07539; Q08029;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Nucleoprotein.
GN NP.
OS Influenza A virus (strain A/Beijing/337/89).
OS Influenza A virus (strain A/Beijing/352/89).
OS Influenza A virus (strain A/Beijing/353/89).
OS Influenza A virus (strain A/Guangdong/338/89).
OS Influenza A virus (strain A/Guangdong/9/87).
OS Influenza A virus (strain A/Memphis/14/85).
OS Influenza A virus (strain A/Memphis/8/88).
OS Influenza A virus (strain A/Shanghai/16/89).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=38951, 38952, 73485, 38961, 38962, 38969, 11441, 38978;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93233239; PubMed-8474171;
RA Shu L.L., Bean W.J., Webster R.G.;
RT "Analysis of the evolution and variation of the human influenza A
RL virus nucleoprotein gene from 1933 to 1990.";
RL J. Virol. 67:2723-2729(1993).
CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M23976; AAA43451.1; -
DR FIRM; D31831; VHI61.
DR InterPro: IPR002141; Flu_NP.
DR Pfam: PF00506; Flu_NP; 1.
KW Nucleoprotein.
FT VARIANT 23 23 N -> T (IN COLD-ADAPTED ISOLATE).
FT VARIANT 34 34 G -> D (IN COLD-ADAPTED ISOLATE).
FT SEQUENCE 498 AA: 55963 MW: E51D08494E9FA075 CRC64;
SQ
Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 DNPQHKTLQ 9
Db 320 ENPAHKSQ 328
RESULT 4
VNUC_IAB39
ID VNUC_IAB39 STANDARD; PRT; 498 AA.
AC Q08028;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Nucleoprotein.
GN NP.
OS Influenza A virus (strain A/Beijing/39/75).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=38955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93233239; PubMed-8474171;
RA Shu L.L., Bean W.J., Webster R.G.;
RT "Analysis of the evolution and variation of the human influenza A
RL virus nucleoprotein gene from 1933 to 1990.";
RL J. Virol. 67:2723-2729(1993).
CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M23976; AAA43451.1; -
DR FIRM; D31831; VHI61.
DR InterPro: IPR002141; Flu_NP.
DR Pfam: PF00506; Flu_NP; 1.
KW Nucleoprotein.
FT VARIANT 23 23 N -> T (IN COLD-ADAPTED ISOLATE).
FT VARIANT 34 34 G -> D (IN COLD-ADAPTED ISOLATE).
FT SEQUENCE 498 AA: 55963 MW: E51D08494E9FA075 CRC64;
SQ
Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 DNPQHKTLQ 9
Db 320 ENPAHKSQ 328

```

QY 1 DNPQHKTLQ 9

Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTLQ 9

```

-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC      EMBL: M30765; AAA43486.1;
CC      InterPro: IPR002141; Flu_NP.
CC      Pfam: PF00506; Flu_NP; 1.
CC      Nucleoprotein.
CC      SEQUENCE 498 AA: 56291 MW: 43990247E22774EB CRC64;
-----
Query Match      72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 DNPQHKQTOL 9
DB      320 ENPAHKSOL 328
      :|||:|||
      :|||:|||

RESULT 7
VNUC_IACAL STANDARD: PRT: 498 AA.
ID VNUC_IACAL STANDARD: PRT: 498 AA.
AC P18069;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Nucleoprotein.
DE Nucleoprotein.
GN NP
OS Influenza A virus (strain A/California/10/78).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OC NCBI_TaxID=11333;
RX MEDLINE=89361370; PubMed=2769232;
RX Altmeiller A., Fitch W.M., Scholtissek C.;
RT "Biological and genetic evolution of the nucleoprotein gene of human
RT Influenza A viruses.";
RL J. Gen. Virol. 70:2111-2119(1989).
CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC      EMBL: D00599; BAA00475.1;
CC      PIR: E36754; VHIVX6.
CC      InterPro: IPR002141; Flu_NP.
CC      Pfam: PF00506; Flu_NP; 1.
CC      Nucleoprotein.
CC      SEQUENCE 498 AA: 55988 MW: 4B09757425A7F8DA CRC64;
-----
Query Match      72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 DNPQHKQTOL 9
DB      320 ENPAHKSOL 328
      :|||:|||
      :|||:|||

RESULT 6
VNUC_IABUD STANDARD: PRT: 498 AA.
ID VNUC_IABUD STANDARD: PRT: 498 AA.
AC P15660;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Nucleoprotein.
DE Nucleoprotein.
GN NP
OS Influenza A virus (strain A/Budgerigar/Hokkaido/1/77).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OC NCBI_TaxID=150159;
RX MEDLINE=90204657; PubMed=2319644;
RX Gorman O.T., Bean W.J., Kawaoka Y., Webster R.G.;
RT "Evolution of the nucleoprotein gene of influenza A virus.";
RL J. Virol. 64:1487-1497(1990).
CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
-----

```

```

RESULT 8
VNUC_IACKG STANDARD; PRT: 498 AA.
AC P16984;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Nucleoprotein.
GN NP.
OS Influenza A virus (strain A/Chicken/Germany/n/49).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11339;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89104698; PubMed=3214270;
RA Reinhardt U., Scholtissek C.;
RT "Comparison of the nucleoprotein genes of a chicken and a mink
in influenza A H10 virus."
RL Arch. Virol. 103:139-145(1988).
CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; M24453; AAA43470.1; -.
DR PIR; A43494; VHIVC1.
DR InterPro; IPR002141; Flu_NP.
DR Pfam; PF00506; Flu_NP; 1.
KW Nucleoprotein.
SQ SEQUENCE 498 AA; 56235 MW; ECEADC066E72FE9 CRC64;

Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHTQL 9
DB 320 ENPAKSQL 328

RESULT 9
VNUC_IACKP STANDARD; PRT: 498 AA.
AC P15661;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Nucleoprotein.
GN NP.
OS Influenza A virus (strain A/Chicken/Pennsylvania/1/83).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11341;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90204657; PubMed=2319644;
RA Gorman O.T., Bean W.J., Kawakita Y., Webster R.G.;
RT "Evolution of the nucleoprotein gene of Influenza A virus."
RL J. Virol. 64:1487-1497(1990).
CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; M30768; AAA43489.1; -.
DR InterPro; IPR002141; Flu_NP.
DR Pfam; PF00506; Flu_NP; 1.
KW Nucleoprotein.
SQ SEQUENCE 498 AA; 56274 MW; HFD13C3A68D848AB CRC64;

Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHTQL 9
DB 320 ENPAKSQL 328

RESULT 10
VNUC_IADAU STANDARD; PRT: 498 AA.
AC P26065;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Nucleoprotein.
GN NP.
OS Influenza A virus (strain A/Duck/Australia/749/80).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11350;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91251223; PubMed=2041090;
RA Gorman O.T., Bean W.J., Kawakita Y., Donatelli I., Guo Y.,
RA Webster R.G.;
RT "Evolution of Influenza A virus nucleoprotein genes: implications for
RT the origins of H1N1 human and classical swine viruses."
RL J. Virol. 65:3704-3714(1991).
CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; M63783; AAA52244.1; -.
DR InterPro; IPR002141; Flu_NP.
DR Pfam; PF00506; Flu_NP; 1.
KW Nucleoprotein.
SQ SEQUENCE 498 AA; 56221 MW; BC4F1473C231C813 CRC64;

Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHTQL 9
DB 320 ENPAKSQL 328

RESULT 11
VNUC_IADBE STANDARD; PRT: 498 AA.
AC P26064;

```

```
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Nucleoprotein.
GN NP.
OS Influenza A virus (strain A/Duck/Beijing/1/78).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11352;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91251223; PubMed=2041090;
RA Gorman O.T., Bean W.J., Kawaoka Y., Donatelli I., Guo Y.,
RA Webster R.G.;
RT "Evolution of influenza A virus nucleoprotein genes: implications for
RT the origins of H1N1 human and classical swine viruses.";
RL J. Virol. 65:3704-3714(1991)
CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63782; AAA52243.1; -
DR InterPro: IPR002141; Flu_NP.
DR Pfam: PF00506; Flu_NP; 1.
KW Nucleoprotein.
SQ SEQUENCE 498 AA; 56253 MW; 49B657952426CEB3 CRC64;

Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTQL 9
DB 320 ENPAHKSQ 328

RESULT 12
VNUC_IADCZ STANDARD; PRT; 498 AA.
AC P15662;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Nucleoprotein.
GN NP.
OS Influenza A virus (strain A/Duck/Czechoslovakia/56).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11353;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90204657; PubMed=2319644;
RA Gorman O.T., Bean W.J., Kawaoka Y., Webster R.G.;
RT "Evolution of the nucleoprotein gene of influenza A virus.";
RL J. Virol. 64:1487-1497(1990).
CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
CC -----
CC EMBL; M30762; AAA43483.1; -
CC InterPro: IPR002141; Flu_NP.
CC Pfam: PF00506; Flu_NP; 1.
KW Nucleoprotein.
SQ SEQUENCE 498 AA; 56244 MW; D393CD14859491ED CRC64;

Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTQL 9
DB 320 ENPAHKSQ 328

RESULT 13
VNUC_IADL STANDARD; PRT; 498 AA.
ID VNUC_IADL
AC P26062;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Nucleoprotein.
GN NP.
OS Influenza A virus (strain A/Duck/England/1/56).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11354;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91251223; PubMed=2041090;
RA Gorman O.T., Bean W.J., Kawaoka Y., Donatelli I., Guo Y.,
RA Webster R.G.;
RT "Evolution of influenza A virus nucleoprotein genes: implications for
RT the origins of H1N1 human and classical swine viruses.";
RL J. Virol. 65:3704-3714(1991).
CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63780; AAA52241.1; -
DR InterPro: IPR002141; Flu_NP.
DR Pfam: PF00506; Flu_NP; 1.
KW Nucleoprotein.
SQ SEQUENCE 498 AA; 5579A6F6DA1934A5 CRC64;

Query Match 72.5%; Score 37; DB 1; Length 498;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTQL 9
DB 320 ENPAHKSQ 328

RESULT 14
VNUC_IADL2 STANDARD; PRT; 498 AA.
ID VNUC_IADL2
AC P26063;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Nucleoprotein.
```

GN NP.
 OS Influenza A virus (strain A/Duck/England/1/62).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza virus A and B group; Influenza A viruses;
 CC Influenza A virus;
 CC NCBI_TaxID=11355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91251223; PubMed=2041090;
 RA Gorman O.T., Bean W.J., Kawakita Y., Donatelli I., Guo Y.,
 RA Webster R.G.;
 RT "Evolution of influenza A virus nucleoprotein genes: implications for
 RT the origins of H1N1 human and classical swine viruses.";
 RL J. Virol. 65:3704-3714(1991).
 CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M63781; AAA52242.1; -
 DR InterPro: IPR002141; Flu_NP.
 DR Pfam: PF00506; Flu_NP; 1.
 KW Nucleoprotein.
 SQ SEQUENCE 498 AA: 56328 MW: 5F0EA230D13C31C0 CRC64;

 Query Match 72.5%; Score 37; DB 1; Length 498;
 Best Local Similarity 66.7%; Pred. No. 8.3;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 QY 1 DNPQHKTQL 9
 :|||:|||
 DB 320 ENPAHKSOL 328

 RESULT 15
 VNUC_IADHK STANDARD; PRT; 498 AA.
 ID VNUC_IADHK
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-MAY-1992 (Rel. 22, Last annotation update)
 DE Nucleoprotein.
 GN NP.
 OS Influenza A virus (strain A/Duck/Hong Kong/7/75).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza virus A and B group; Influenza A viruses;
 CC Influenza A virus.
 CC NCBI_TaxID=11364;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89243210; PubMed=2718389;
 RA Gammel M., Mandler J., Scholtissek C.;
 RT "Two subtypes of nucleoproteins (NP) of influenza A viruses.";
 RL Virology 170:71-80(1989).
 CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M22573; AAA43097.1; -
 DR InterPro: IPR002141; Flu_NP.
 DR PIR: B31470; VHIWN2.

DR Pfam: PF00506; Flu_NP; 1.
 KW Nucleoprotein.
 SQ SEQUENCE 498 AA: 56239 MW: 9C37ACCA466647C7D CRC64;

 Query Match 72.5%; Score 37; DB 1; Length 498;
 Best Local Similarity 66.7%; Pred. No. 8.3;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 QY 1 DNPQHKTQL 9
 :|||:|||
 DB 320 ENPAHKSOL 328

 Search completed: August 6, 2002, 17:07:36
 Job time: 878 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 17:09:32 : Search time 111.35 Seconds
(without alignments)
13.983 Million cell updates/sec

Title: US-10-020-139-2_COPY_237_245

Perfect score: 51

Sequence: 1 DNPQHKTL 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_minc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_todent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	51	100.0	249	Q9BQ00	Q9BQ00 homo sapien
2	51	100.0	249	Q96DR5	Q96DR5 homo sapien
3	39	76.5	252	O07163	O07163 mycoplasma
4	37	72.5	121	O91U38	O91U38 influenza a
5	37	72.5	152	O91IG3	O91IG3 influenza a
6	37	72.5	152	O91IG2	O91IG2 influenza a
7	37	72.5	152	O91IG1	O91IG1 influenza a
8	37	72.5	152	O91IG0	O91IG0 influenza a
9	37	72.5	152	O91IF9	O91IF9 influenza a
10	37	72.5	152	O91IF8	O91IF8 influenza a
11	37	72.5	152	O91IF7	O91IF7 influenza a
12	37	72.5	152	O91IF6	O91IF6 influenza a
13	37	72.5	152	O91IF5	O91IF5 influenza a
14	37	72.5	152	O91IF4	O91IF4 influenza a
15	37	72.5	152	O91IF3	O91IF3 influenza a
16	37	72.5	152	O91IF2	O91IF2 influenza a

17	37	72.5	152	12	Q91IF1	Q91IF1 influenza a
18	37	72.5	152	12	Q91IF0	Q91IF0 influenza a
19	37	72.5	152	12	Q91IE9	Q91IE9 influenza a
20	37	72.5	152	12	Q91IE8	Q91IE8 influenza a
21	37	72.5	152	12	Q91IE7	Q91IE7 influenza a
22	37	72.5	152	12	Q91IE6	Q91IE6 influenza a
23	37	72.5	152	12	Q91IE5	Q91IE5 influenza a
24	37	72.5	152	12	Q91IE4	Q91IE4 influenza a
25	37	72.5	152	12	Q91IE3	Q91IE3 influenza a
26	37	72.5	152	12	Q91IE2	Q91IE2 influenza a
27	37	72.5	152	12	Q91IE1	Q91IE1 influenza a
28	37	72.5	152	12	Q91IE0	Q91IE0 influenza a
29	37	72.5	152	12	Q91ID9	Q91ID9 influenza a
30	37	72.5	152	12	Q91ID8	Q91ID8 influenza a
31	37	72.5	152	12	Q91ID7	Q91ID7 influenza a
32	37	72.5	152	12	Q91ID6	Q91ID6 influenza a
33	37	72.5	152	12	Q91ID5	Q91ID5 influenza a
34	37	72.5	152	12	Q91ID4	Q91ID4 influenza a
35	37	72.5	152	12	Q91ID3	Q91ID3 influenza a
36	37	72.5	152	12	Q91ID2	Q91ID2 influenza a
37	37	72.5	152	12	Q91ID1	Q91ID1 influenza a
38	37	72.5	152	12	Q91ID0	Q91ID0 influenza a
39	37	72.5	152	12	Q91IC9	Q91IC9 influenza a
40	37	72.5	152	12	Q91IC8	Q91IC8 influenza a
41	37	72.5	152	12	Q91IC7	Q91IC7 influenza a
42	37	72.5	152	12	Q91IC6	Q91IC6 influenza a
43	37	72.5	152	12	Q91IC5	Q91IC5 influenza a
44	37	72.5	152	12	Q91IC4	Q91IC4 influenza a
45	37	72.5	152	12	Q91IC3	Q91IC3 influenza a

ALIGNMENTS

RESULT 1

Q9BQ00 PRELIMINARY; PRT; 249 AA.

AC Q9BQ00;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE BA49G10.1 (SIMILAR TO BOVINE SALIVARY PROTEIN BSP30).

GN BA49G10.1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Tracey A.;

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL121901; CAC03546.1; ..

SQ SEQUENCE 249 AA; 27011 MW; E64E0794A1B4DB7D CRC64;

Query Match 100.0%; Score 51; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNPQHKTL 9
|||||||

Db 237 DNPQHKTL 245

RESULT 2

Q96DR5 PRELIMINARY; PRT; 249 AA.

ID Q96DR5;

AC Q96DR5;

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE PAROTID SECRETORY PROTEIN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PAROTID;
 RA Venkatesh S.G., Geetha C., Gorr S.-U.;
 RT "A member of the PSP/plunc family of bpr proteins is expressed in the
 human parotid gland."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF432917; AAL28113.1;
 SQ SEQUENCE 249 AA; 27110 MW; FD54B624A1A4CA7C CRC64;

Query Match 100.0%; Score 51; DR 4; Length 249;
 Best Local Similarity 100.0%; Pred. No. 0.03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNPQHKTKQL 9
 |||||
 Db 237 DNPQHKTKQL 245

RESULT 3
 O07163 PRELIMINARY; PRT: 252 AA.
 AC O07163;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last annotation update)
 DE PC142-56 IMMUNODOMINANT MEMBRANE PROTEIN (FRAGMENT).
 GN MS2/28.
 OS Mycoplasma synoviae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2109;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VWU 1853;
 RA ben Abdelmoumen B., Roy R.S., Brousseau R.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U66315; AAB58471.1;
 FT NON-TER 1
 SQ SEQUENCE 252 AA; 27323 MW; 262E4302C6B5226 CRC64;

Query Match 76.5%; Score 39; DR 2; Length 252;
 Best Local Similarity 87.5%; Pred. No. 7.2;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 NPOHKTQL 9
 |||||
 Db 230 NPOHKTQL 237

RESULT 4
 Q91038 PRELIMINARY; PRT: 121 AA.
 AC Q91038;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE NUCLEOCAPSID PROTEIN (FRAGMENT).
 GN NP.
 OS Influenza A virus (A/Hong Kong/507/97(H5N1)).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
 OX NCBI_TaxID=153225;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/HONG KONG/507/97;
 RA Shaw M.W., Cooper L.A., Xu X., Thompson W.W., Krauss S.L., Guan Y.,
 Zhou N.N., Klimov A., Cox N.J., Webster R.G., Lim W., Shortridge K.F.,

RA Subbarao K.;
 RT "Avian Influenza A H5N1 and H9N2 viruses bearing a specific
 constellation of nonglycoprotein genes caused illness in humans."
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF255763; AAK49289.1;
 FT NON-TER 1
 FT NON-TER 121
 SQ SEQUENCE 121 AA; 13295 MW; 2DEE1A250286E685 CRC64;

Query Match 72.5%; Score 37; DR 12; Length 121;
 Best Local Similarity 66.7%; Pred. No. 8.4;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTKQL 9
 :|||:
 Db 58 ENPAHKSQ 66

RESULT 5
 Q911G3 PRELIMINARY; PRT: 152 AA.
 AC Q911G3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE NUCLEOPROTEIN (FRAGMENT).
 GN NP.
 OS Influenza A virus (A/Netherlands/001/94 (H3N2)).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
 OX NCBI_TaxID=132728;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20347342; PubMed=10888619;
 RA Voeten J.T., Besterbroer T.M., Nieuwkoop N.J., Fouchier R.A.,
 Osterhaus A.D., Rimmelzwaan G.F.;
 RT "Antigenic drift in the influenza A virus (H3N2) nucleoprotein and
 escape from recognition by cytotoxic T lymphocytes."
 RL J. Virol. 74:6800-6807(2000).
 DR EMBL: AF225709; AAF87330.1;
 DR InterPro: IPR002141; Flu_NP.
 DR Pfam: PF00506; Flu_NP; 1.
 FT NON-TER 152
 FT NON-TER 152
 SQ SEQUENCE 152 AA; 16846 MW; F9F9594CE8F2D01A CRC64;

Query Match 72.5%; Score 37; DR 12; Length 152;
 Best Local Similarity 66.7%; Pred. No. 11;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTKQL 9
 :|||:
 Db 81 ENPAHKSQ 89

RESULT 6
 Q911G2 PRELIMINARY; PRT: 152 AA.
 AC Q911G2;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE NUCLEOPROTEIN (FRAGMENT).
 GN NP.
 OS Influenza A virus (A/Netherlands/002/94 (H3N2)).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
 OX NCBI_TaxID=132729;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20347342; PubMed=10888619;

```

RA Voeten J.T., Bestebroer T.M., Nieuwkoop N.J., Fouchier R.A.,
RA Osterhaus A.D., Rimmelzwaan G.F.;
RT "Antigenic drift in the Influenza A virus (H3N2) nucleoprotein and
RT escape from recognition by cytotoxic T lymphocytes.";
RL J. Virol. 74:6800-6807(2000).
DR EMBL; AF225710; AAF87331.1; -.
DR InterPro: IPR002141; Flu_NP.
DR Pfam: PF00506; Flu_NP; 1.
FT NON_TER 152
FT NON_TER 152
SQ SEQUENCE 152 AA; 16846 MW; F9F9594CE8F2D01A CRC64;

Query Match 72.5%; Score 37; DB 12; Length 152;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTL 9
Db 81 ENPAKSQL 89

RESULT 7
ID Q9IIG1 PRELIMINARY; PRT; 152 AA.
AC Q9IIG1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NUCLEOPROTEIN (FRAGMENT).
GN NP.
OS Influenza A virus (A/Netherlands/005/94 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX NCBI_TaxID=132730;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347342; PubMed=10888619;
RA Voeten J.T., Bestebroer T.M., Nieuwkoop N.J., Fouchier R.A.,
RA Osterhaus A.D., Rimmelzwaan G.F.;
RT "Antigenic drift in the Influenza A virus (H3N2) nucleoprotein and
RT escape from recognition by cytotoxic T lymphocytes.";
RL J. Virol. 74:6800-6807(2000).
DR EMBL; AF225711; AAF87332.1; -.
DR InterPro: IPR002141; Flu_NP.
DR Pfam: PF00506; Flu_NP; 1.
FT NON_TER 152
FT NON_TER 152
SQ SEQUENCE 152 AA; 16846 MW; F9F9594CE8F2D01A CRC64;

Query Match 72.5%; Score 37; DB 12; Length 152;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTL 9
Db 81 ENPAKSQL 89

RESULT 8
ID Q9IIG0 PRELIMINARY; PRT; 152 AA.
AC Q9IIG0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NUCLEOPROTEIN (FRAGMENT).
GN NP.
OS Influenza A virus (A/Netherlands/007/94 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX NCBI_TaxID=132731;

```

```

[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347342; PubMed=10888619;
RA Voeten J.T., Bestebroer T.M., Nieuwkoop N.J., Fouchier R.A.,
RA Osterhaus A.D., Rimmelzwaan G.F.;
RT "Antigenic drift in the Influenza A virus (H3N2) nucleoprotein and
RT escape from recognition by cytotoxic T lymphocytes.";
RL J. Virol. 74:6800-6807(2000).
DR EMBL; AF225712; AAF87333.1; -.
DR InterPro: IPR002141; Flu_NP.
DR Pfam: PF00506; Flu_NP; 1.
FT NON_TER 152
FT NON_TER 152
SQ SEQUENCE 152 AA; 16846 MW; F9F9594CE8F2D01A CRC64;

Query Match 72.5%; Score 37; DB 12; Length 152;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTL 9
Db 81 ENPAKSQL 89

RESULT 9
ID Q9IIF9 PRELIMINARY; PRT; 152 AA.
AC Q9IIF9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NUCLEOPROTEIN (FRAGMENT).
GN NP.
OS Influenza A virus (A/Netherlands/008/94 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX NCBI_TaxID=132732;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347342; PubMed=10888619;
RA Voeten J.T., Bestebroer T.M., Nieuwkoop N.J., Fouchier R.A.,
RA Osterhaus A.D., Rimmelzwaan G.F.;
RT "Antigenic drift in the Influenza A virus (H3N2) nucleoprotein and
RT escape from recognition by cytotoxic T lymphocytes.";
RL J. Virol. 74:6800-6807(2000).
DR EMBL; AF225713; AAF87334.1; -.
DR InterPro: IPR002141; Flu_NP.
DR Pfam: PF00506; Flu_NP; 1.
FT NON_TER 152
FT NON_TER 152
SQ SEQUENCE 152 AA; 16846 MW; F9F9594CE8F2D01A CRC64;

Query Match 72.5%; Score 37; DB 12; Length 152;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTL 9
Db 81 ENPAKSQL 89

RESULT 10
ID Q9IIF8 PRELIMINARY; PRT; 152 AA.
AC Q9IIF8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NUCLEOPROTEIN (FRAGMENT).
GN NP.
OS Influenza A virus (A/Netherlands/017/94 (H3N2)).

```

```

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX NCBI_TaxID=132733;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347342; PubMed=10888619;
RA Voeten J.T., Bestebroer T.M., Nieuwkoop N.J., Fouchier R.A.,
RA Osterhaus A.D., Rimmelzwaan G.F.;
RT "Antigenic drift in the influenza A virus (H3N2) nucleoprotein and
RT escape from recognition by cytotoxic T lymphocytes.";
RL J. Virol. 74:6800-6807(2000).
DR EMBL: AF225714; AAF87336.1;
DR InterPro: IPR002141; Flu_NP.
DR Pfam: PF00506; Flu_NP; 1.
DR NON_TER 1
FT NON_TER 152
FT SEQUENCE 152 AA; 16846 MW; F9F9594CE8F2D01A CRC64;

Query Match 72.5%; Score 37; DB 12; Length 152;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKQTQL 9
Db 81 ENPAHKSQ 89

RESULT 11
Q9IIF7 PRELIMINARY; PRT; 152 AA.
AC Q9IIF7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NUCLEOPROTEIN (FRAGMENT).
GN NP.
OS Influenza A virus (A/Netherlands/018/94 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX NCBI_TaxID=132734;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347342; PubMed=10888619;
RA Voeten J.T., Bestebroer T.M., Nieuwkoop N.J., Fouchier R.A.,
RA Osterhaus A.D., Rimmelzwaan G.F.;
RT "Antigenic drift in the influenza A virus (H3N2) nucleoprotein and
RT escape from recognition by cytotoxic T lymphocytes.";
RL J. Virol. 74:6800-6807(2000).
DR EMBL: AF225715; AAF87336.1;
DR InterPro: IPR002141; Flu_NP.
DR Pfam: PF00506; Flu_NP; 1.
DR NON_TER 1
FT NON_TER 152
FT SEQUENCE 152 AA; 16846 MW; F9F9594CE8F2D01A CRC64;

Query Match 72.5%; Score 37; DB 12; Length 152;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKQTQL 9
Db 81 ENPAHKSQ 89

RESULT 12
Q9IIF6 PRELIMINARY; PRT; 152 AA.
AC Q9IIF6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

```

```

DE NUCLEOPROTEIN (FRAGMENT).
GN NP.
OC Influenza A virus (A/Netherlands/020/94 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX NCBI_TaxID=132735;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347342; PubMed=10888619;
RA Voeten J.T., Bestebroer T.M., Nieuwkoop N.J., Fouchier R.A.,
RA Osterhaus A.D., Rimmelzwaan G.F.;
RT "Antigenic drift in the influenza A virus (H3N2) nucleoprotein and
RT escape from recognition by cytotoxic T lymphocytes.";
RL J. Virol. 74:6800-6807(2000).
DR EMBL: AF225716; AAF87337.1;
DR InterPro: IPR002141; Flu_NP.
DR Pfam: PF00506; Flu_NP; 1.
DR NON_TER 1
FT NON_TER 152
FT SEQUENCE 152 AA; 16846 MW; F9F9594CE8F2D01A CRC64;

Query Match 72.5%; Score 37; DB 12; Length 152;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKQTQL 9
Db 81 ENPAHKSQ 89

RESULT 13
Q9IIF5 PRELIMINARY; PRT; 152 AA.
AC Q9IIF5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NUCLEOPROTEIN (FRAGMENT).
GN NP.
OS Influenza A virus (A/Netherlands/021/94 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX NCBI_TaxID=132736;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347342; PubMed=10888619;
RA Voeten J.T., Bestebroer T.M., Nieuwkoop N.J., Fouchier R.A.,
RA Osterhaus A.D., Rimmelzwaan G.F.;
RT "Antigenic drift in the influenza A virus (H3N2) nucleoprotein and
RT escape from recognition by cytotoxic T lymphocytes.";
RL J. Virol. 74:6800-6807(2000).
DR EMBL: AF225717; AAF87338.1;
DR InterPro: IPR002141; Flu_NP.
DR Pfam: PF00506; Flu_NP; 1.
DR NON_TER 1
FT NON_TER 152
FT SEQUENCE 152 AA; 16846 MW; F9F9594CE8F2D01A CRC64;

Query Match 72.5%; Score 37; DB 12; Length 152;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKQTQL 9
Db 81 ENPAHKSQ 89

RESULT 14
Q9IIF4 PRELIMINARY; PRT; 152 AA.
AC Q9IIF4;

```

Search completed: August 6, 2002, 17:09:32
Job time: 934 sec

DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NUCLEOPROTEIN (FRAGMENT).
GN NP.
OS Influenza A virus (A/Netherlands/022/94 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX NCBI_TaxID=132737;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347342; PubMed=10888619;
RA Voeten J.T., Bestebroer T.M., Nieuwkoop N.J., Fouchier R.A.,
RA Osterhaus A.D., Rimmelzwaan G.F.;
RT "Antigenic drift in the Influenza A virus (H3N2) nucleoprotein and
RL J. Virol. 74:6800-6807(2000).
DR EMBL; AF225718; AAF87340.1;
DR InterPro; IPR002141; Flu_NP.
DR Pfam; PF00506; Flu_NP; 1.
FT NON_TER 1
FT NON_TER 152
SQ SEQUENCE 152 AA; 16846 MW; F9F9594CE8F2D01A CRC64;

Query Match 72.5%; Score 37; DB 12; Length 152;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPOHKTOL 9
DB 81 ENPAKSQL 89

RESULT 15

QY 09IIF3 PRELIMINARY; PRT: 152 AA.
AC 09IIF3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NUCLEOPROTEIN (FRAGMENT).
GN NP.
OS Influenza A virus (A/Netherlands/026/94 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX NCBI_TaxID=132738;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347342; PubMed=10888619;
RA Voeten J.T., Bestebroer T.M., Nieuwkoop N.J., Fouchier R.A.,
RA Osterhaus A.D., Rimmelzwaan G.F.;
RT "Antigenic drift in the Influenza A virus (H3N2) nucleoprotein and
RL J. Virol. 74:6800-6807(2000).
DR EMBL; AF225719; AAF87340.1;
DR InterPro; IPR002141; Flu_NP.
DR Pfam; PF00506; Flu_NP; 1.
FT NON_TER 1
FT NON_TER 152
SQ SEQUENCE 152 AA; 16846 MW; F9F9594CE8F2D01A CRC64;

Query Match 72.5%; Score 37; DB 12; Length 152;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPOHKTOL 9
DB 81 ENPAKSQL 89

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 16:52:54 ; Search time 138.55 Seconds
(without alignments)
7.215 Million cell updates/sec

Title: US-10-020-139-2_COPY_237_245

Perfect score: 51

Sequence: 1 DNPQKHTOL 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802:*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	100.0	249	19	AAW69221
2	51	100.0	249	19	AAW60682
3	51	100.0	249	21	AAW24069
4	51	100.0	249	21	AAW23765
5	51	100.0	249	22	AAW75351
6	51	100.0	260	22	AAW25745
7	37	72.5	217	22	AAE09054
8	37	72.5	225	22	AAE09053
9	37	72.5	498	19	AAW68409
10	37	72.5	498	19	AAW68407
11	37	72.5	499	19	AAW55986

12	37	72.5	917	14	AAW36821	PE binding/translo
13	37	72.5	917	14	AAW32469	PE binding and tra
14	36	70.6	104	21	AAW15944	E. coli proliferat
15	36	70.6	132	20	AAW29160	Amino acid sequenc
16	36	70.6	817	22	ABW70027	Drosophila melanog
17	35	68.6	496	22	ABW52510	Escherichia coli p
18	34	66.7	70	22	AAU62675	Propionibacterium
19	34	66.7	213	22	AAE09058	Equine influenza v
20	34	66.7	217	22	AAE09056	Equine influenza v
21	34	66.7	221	22	AAE09057	Equine influenza v
22	34	66.7	225	22	AAE09055	Equine influenza v
23	34	66.7	239	21	AAW41169	Human ORF4 ORF933
24	34	66.7	363	22	ABW69159	Drosophila melanog
25	34	66.7	412	14	AAW38309	Sequence of the P4
26	34	66.7	1063	22	ABW12679	Novel human diagno
27	33	64.7	63	22	AAU42527	Propionibacterium
28	33	64.7	87	22	AAU49827	Propionibacterium
29	33	64.7	306	21	AAW09000	Arabidopsis thalia
30	33	64.7	306	21	AAW42634	Arabidopsis thalia
31	33	64.7	345	19	AAW41166	Metal-regulated tr
32	33	64.7	362	20	AAW38750	Neisseria gonorrhoe
33	33	64.7	371	20	AAW38751	Neisseria gonorrhoe
34	33	64.7	371	20	AAW38748	Neisseria meningit
35	33	64.7	371	20	AAW08963	A. gossypii ORF 2
36	33	64.7	393	22	AAW5761	Human protein sequ
37	33	64.7	409	21	AAW51064	S. tendae nicomyci
38	33	64.7	409	21	AAW50820	Streptomyces tende
39	33	64.7	460	22	AAW81272	Human AFP protein
40	33	64.7	587	22	ABW60969	Drosophila melanog
41	33	64.7	874	22	ABW62601	Drosophila melanog
42	33	64.7	1001	13	AAW325068	hLIF-R. Homo sapi
43	32	62.7	33	22	ABW23036	Protein #5035 enco
44	32	62.7	85	19	AAW38322	S. pneumoniae glut
45	32	62.7	98	19	AAW60943	Streptococcus pneu

ALIGNMENTS

RESULT 1

AAW69221
ID AAW69221 standard; Protein; 249 AA.

XX AC AAW69221;

XX DT 16-OCT-1998 (first entry)

XX DE Human parotid secretory protein.

XX KW Parotid secretory protein; HPSP; digestive disorder; endocrine disorder;
non-immune defensive disorder; immune system disorder; cancer; human;
therapy; diagnosis.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..18 /note= "signal peptide"

FT Protein 19..249

FT FT /note= "mature hPSP"

XX PN W09828420-A1.

XX PD 02-JUL-1998.

XX PF 18-DEC-1997; 97WO-US23522.

XX PR 23-DEC-1996; 96US-0034429.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Duan R, Ruben SM;

XX XX

DR WPI: 1998-377651/32.
 DR N-PSDB; AAV44759.
 XX
 PT New nucleic acid encoding human parotid secretory protein or its
 fragments - useful for diagnosis and treatment of, e.g. digestive
 PT and endocrine disorders and for drug screening
 XX
 PS Claim 16; Fig 1; 94pp; English.
 XX
 CC This sequence is the human parotid secretory protein (hPSP) of the
 CC invention. The hPSP DNA is useful for chromosome identification and
 CC isolation of the corresponding genomic DNA. The DNA and protein can be
 CC used to detect abnormal levels of hPSP (in standard blotting,
 CC amplification or immuno assays), particularly for diagnosis of digestive,
 CC non-immune defensive, endocrine or immune system disorders. A particular
 CC application is diagnosis of cancers of the salivary gland, thymus and
 CC pancreas which are associated with high levels of hPSP. The protein is
 CC also useful as antifungal, antibacterial, antiparasitic and antiviral
 CC agents and may be expressed in vivo from the DNA. The protein, or cells
 CC expressing it, are used in screening tests to identify specific
 CC (antagonists, e.g. antibodies (Ab), polypeptides and antisense nucleic
 CC acids, which are potentially useful for treating conditions associated
 CC with excessive hPSP production. Cells containing the DNA are used to
 CC express the recombinant protein and this can be used to raise Ab, useful
 CC for diagnosis, therapy, for affinity purification and to identify
 CC hPSP-binding proteins.
 XX
 SQ Sequence 249 AA;

Query Match 100.0%; Score 51; DB 19; Length 249;
 Best Local Similarity 100.0%; Pred. No. 0.091;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DNPOHKTQL 9
 Db 237 dnpqhtql 245
 |||||

RESULT 2
 AAW60682
 ID AAW60682 standard; Protein: 249 AA.
 AC AAW60682;
 DT 18-SEP-1998 (first entry)
 DE Human parotid secretory protein (hPSP).
 XX Parotid secretory protein; human; cancer; autoimmune disease;
 KW secretory tissue; gastrointestinal tissue; hPSP; Sjogren's syndrome;
 KW Graves disease; thyroiditis; insulin-dependent diabetes; pancreatitis;
 KW ulcerative colitis; Crohn's disease; atrophic gastritis.
 XX
 OS Homo sapiens.
 XX
 PN WO9821329-A1.
 XX
 PD 22-MAY-1998.
 XX
 PE 07-NOV-1997; 97WO-US20651.
 XX
 PR 14-NOV-1996; 96US-0749288.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Bandman O, Goll SK;
 XX
 DR WPI: 1998-297933/26.
 DR N-PSDB; AAV37699.
 XX
 PT New parotid secretory protein - useful for, e.g. treatment of cancer
 PT and autoimmune disease, particularly of secretory or

PT gastrointestinal tissues
 XX
 PS Claim 1; Fig 1A-C; 65pp; English.
 XX
 CC This represents a human parotid secretory protein (hPSP). Antagonists
 CC that bind specifically to, and modulate activity of hPSP are used to
 CC treat cancer and autoimmune diseases particularly of secretory or
 CC gastrointestinal tissue, e.g. cancer of salivary gland, thyroid,
 CC prostate, breast, gastrointestinal tract or pancreas, Sjogren's syndrome,
 CC Graves disease, thyroiditis, insulin-dependent diabetes, pancreatitis,
 CC ulcerative colitis, Crohn's disease and atrophic gastritis. Host cells
 CC containing expression vectors comprising the hPSP nucleic acid are used
 CC to produce recombinant hPSP which is used to generate antibodies and to
 CC screen for its antagonists. Antibodies are useful directly as
 CC antagonists, to transport drugs to hPSP-expressing cells, to detect cells
 CC that express hPSP to monitor patients being treated with hPSP, and for
 CC purification of hPSP from natural sources. Expression of hPSP may be
 CC indicate cell proliferation. hPSP nucleic acid or its fragments are used
 CC to detect hPSP-encoding sequences (optionally after amplification by PCR)
 CC by hybridisation, particularly for diagnosis and monitoring of disease,
 CC but also for mapping the chromosomal sequence.
 XX
 SQ Sequence 249 AA;

Query Match 100.0%; Score 51; DB 19; Length 249;
 Best Local Similarity 100.0%; Pred. No. 0.091;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DNPOHKTQL 9
 Db 237 dnpqhtql 245
 |||||

RESULT 3
 AAB24069
 ID AAB24069 standard; Protein: 249 AA.
 AC AAB24069;
 DT 29-JAN-2001 (first entry)
 DE Human PRO1025 protein sequence SEQ ID NO:38.
 XX
 KW Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
 KW proliferation; tumorigenesis; identification; cancer; cytostatic;
 KW neutropenic; neutrophilic; antineoplastic; immunosuppressive;
 KW immunostimulant; antineoplastic; leukaemia; lymphoid malignancy;
 KW neuronal disorder; glioma disorder; astrocytoma disorder; angiogenic;
 KW hypothalamic disorder; glandular disorder; macropapillary disorder;
 KW epithelial disorder; stromal disorder; blastocoele disorder;
 KW inflammatory disorder; immunologic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200053755-A2.
 XX
 PD 14-SEP-2000.
 XX
 PE 06-JAN-2000; 2000WO-US00376.
 XX
 PR 08-MAR-1999; 99WO-US05028.
 PR 02-JUN-1999; 99WO-US12252.
 PR 23-JUN-1999; 99US-0141037.
 PR 07-JUL-1999; 99US-0143048.
 PR 26-JUL-1999; 99US-0145698.
 PR 30-NOV-1999; 99WO-US28313.
 PR 20-DEC-1999; 99WO-US30911.
 PR 05-JAN-2000; 2000WO-US00219.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;

PI Watanabe CK, Wood WI;
 XX WPI: 2000-572270/53.
 DR N-PSDB: AAC58379.
 XX
 XX Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
 PT treatment, diagnosis and prevention of cancer.
 PT
 PS Claim 61: Fig 26: 286pp; English.
 XX
 XX The present invention describes an isolated antibody that binds to
 CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
 CC PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
 CC PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
 CC PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094, cell
 CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
 CC growth. The PRO polypeptides and nucleotides are useful in the
 CC treatment, diagnosis and prevention of cancer. The antibodies and other
 CC anti-tumour compounds may be used to treat various conditions, including
 CC those characterised by overexpression and/or activation of the amplified
 CC PRO genes. Exemplary conditions or disorders to be treated with such
 CC antibodies and other compounds include benign or malignant tumours
 CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
 CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
 CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
 CC leukaemias and lymphoid malignancies, other disorders such as neuronal,
 CC glial, astrocytic, hypothalamic and other glandular, macrophagal,
 CC epithelial, stromal and blastocoeleic disorders, and inflammatory,
 CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
 CC primers and hybridisation probes used in the isolation of the human PRO
 CC sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
 CC PRO polynucleotide and protein sequences given in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 249 AA;

Query Match 100.0%; Score 51; DB 21; Length 249;
 Best Local Similarity 100.0%; Pred. No. 0.091;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DNPOHKTQL 9
 |||||
 Db 237 dnpqhtktql 245

RESULT 4
 AAB25765
 ID AAB25765 standard; protein: 249 AA.

XX AAB25765;

XX 28-NOV-2000 (first entry)

XX Human secreted protein SEQ ID #77.

XX Human; secreted protein; forensic procedure; gene therapy;
 KW chromosome mapping; cancer; autoimmune disease; cardiovascular disorder;
 KW cystic fibrosis; hypothyroidism; immunological disorder; amyloidosis;
 KW brain disorder; skeletal muscle disorder; eye disorder; obesity;
 KW mitochondrial cytopathy; diabetes; atherosclerosis; Alzheimer's disease;
 KW neurodegenerative disorder; graft rejection; dementia; hyperlipidaemia;
 KW septic shock; impotence.

XX Homo sapiens.

XX WO200037491-A2.

XX 29-JUN-2000.

XX 20-DEC-1999; 99WO-IB02058.

XX 22-DEC-1998; 98US-0113686.

PR 25-JUN-1999; 99US-0141032.
 XX (GEST) GENSET.
 XX Bouqueleret L, Dumas J, Duclert A;
 XX WPI: 2000-442637/38.
 DR N-PSDB: AAA87727.
 XX
 XX Polynucleotides and polypeptides encoding proteins with signal
 PT peptides, useful in diagnostic, forensic, gene therapy and chromosome
 PT mapping procedures.
 XX
 XX Claim 9: Figure 10; 306pp; English.
 XX
 XX This sequence represents a human secreted protein amino acid sequence.
 CC The invention relates to sequences AAA87725-A87774 which encode human
 CC secreted proteins AAB25763-B25812. The proteins include signal peptides.
 CC Included in the invention are a host cell containing one of the cDNA
 CC sequences, and a purified antibody capable of binding to one of the
 CC secreted proteins. Also contained in the invention are methods for
 CC storing the sequence data on a computer system, and a method for
 CC identifying features of the cDNA sequences using a computer programme.
 CC The cDNAs are useful for expressing secreted proteins or fragments to
 CC obtain antibodies capable of specifically binding to the secreted
 CC proteins. The cDNAs may also be useful in diagnostic, forensic, gene
 CC therapy and chromosome mapping procedures and may be used to design
 CC expression vectors and secretion vectors. The proteins of the invention
 CC may be used to treat diseases including cancer, autoimmune diseases,
 CC cardiovascular disorders, cystic fibrosis, hypothyroidism, immunological
 CC disorders, amyloidosis, brain disorders, skeletal muscle disorders, eye
 CC disorders, obesity, mitochondrial cytopathies, diabetes, atherosclerosis,
 CC neurodegenerative disorders, graft rejection, Alzheimer's disease,
 CC dementia, hyperlipidaemia, septic shock and impotence.
 XX
 SQ Sequence 249 AA;

Query Match 100.0%; Score 51; DB 21; Length 249;
 Best Local Similarity 100.0%; Pred. No. 0.091;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DNPOHKTQL 9
 |||||
 Db 237 dnpqhtktql 245

RESULT 5

AAAB75351
 ID AAAB75351 standard; protein: 249 AA.

XX AAAB75351;

XX 05-APR-2001 (first entry)

XX Human secreted protein #10.

XX Secreted protein; prevention; treatment; diagnosis; disease;
 KW infection.

XX Homo sapiens.

XX WO200100806-A2.

XX 04-JAN-2001.

XX 21-JUN-2000; 2000WO-IB00951.

XX 25-JUN-1999; 99US-0141032.

XX 21-DEC-1999; 99US-0469099.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Bougueleret L, Jobert S;
 DR WPI; 2001-071487/08.

XX 49 Secreted proteins and the nucleic acids encoding them, useful in
 PT gene therapy and for detecting similar sequences in samples -

XX Claim 10; Page 281; 307pp; English.

XX The present invention relates to 49 secreted proteins and the cDNAs
 CC encoding them. The protein and nucleic acids may be used in the
 CC prevention, treatment and diagnosis of diseases associated with
 CC inappropriate protein expression.

XX Sequence 249 AA;

Query Match 100.0%; Score 51; DB 22; Length 249;
 Best Local Similarity 100.0%; Pred. No. 0.091; 0; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0

Qy 1 DNPOHKTOL 9
 |||||
 Db 237 dnpqhtql 245

RESULT 6

AA025745
 ID AA025745 standard; Protein; 260 AA.

XX AA025745;

XX 16-OCT-2001 (first entry)

XX Human protein sequence SEQ ID NO:1260.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.

XX Homo sapiens.

XX WO200153455-A2.

XX 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 9905-0471275.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457603/49.

DR N-PSDB; AA099686.

XX Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX Claim 20; Page 260; 1217pp; English.

XX AA099166 to AA099904 encode the human proteins given in AA025225 to
 CC AA025963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
 CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.

XX Sequence 260 AA;

Query Match 100.0%; Score 51; DB 22; Length 260;
 Best Local Similarity 100.0%; Pred. No. 0.095;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DNPOHKTOL 9
 |||||
 Db 248 dnpqhtql 256

RESULT 7

AA09054

ID AA09054 standard; Protein; 217 AA.

XX AA09054;

XX 15-NOV-2001 (first entry)

XX Equine influenza virus H3N8 PeitlNP-C-217 protein.

XX Equine influenza virus; ei; cold adaptation; temperature sensitivity;
 KW vaccine; newtNP-C-656 DNA; PeitlNP-C-217 protein.

XX Equine influenza virus H3N8.

XX WO200160849-A2.

XX 23-AUG-2001.

XX 16-FEB-2001; 2001WO-US05048.

XX 16-FEB-2000; 2000US-0506286.

XX (UYPI-) UNIV PITTSBURGH.

XX Dowling PW, Youngner JS;

XX WPI; 2001-522584/57.

DR N-PSDB; AAD15725.

XX Novel isolated equine influenza virus (wild-type and cold-adapted)
 PT proteins and viruses containing nucleic acid molecules encoding the
 PT proteins, which are useful for protecting animals from influenza virus
 PT infections -

XX Claim 5; Page 166-167; 172pp; English.

XX The patent discloses cold-adapted equine influenza viruses and
 CC reassortant influenza A viruses comprising at least one genome
 CC segment of such an equine influenza virus, wherein the equine
 CC influenza virus genome segment confers at least one identifying
 CC phenotype of the cold-adapted equine influenza virus, such as
 CC cold adaptation, temperature sensitivity, dominant interference
 CC or attenuation. The viruses are useful for protecting animals
 CC from diseases caused by influenza viruses. They are also used
 CC as vaccines. The present sequence is equine influenza (ei) virus
 CC H3N8 Peiwt1 (wild type) NP-C-217 protein which is encoded by
 CC neiwtNP-C-656 DNA.

XX Sequence 217 AA:

Query Match 72.5%; Score 37; DB 22; Length 217;
 Best Local Similarity 66.7%; Pred. No. 34;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 DNPOHKTQL 9
 :|||:|||
 Db 39 enpahksq 47

RESULT 8
 AAE09053
 ID AAE09053 standard; Protein; 225 AA.
 XX
 AC AAE09053;
 XX
 DT 15-NOV-2001 (first entry)
 XX
 DE Equine influenza virus H3N8 PeiwtNP-C-225 protein.
 XX
 KW Equine influenza virus; ei; cold adaptation; temperature sensitivity;
 KW vaccine; neiwtNP-C-679 DNA; PeiwtNP-C-225 protein.
 XX
 OS Equine influenza virus H3N8.

XX Key Location/Qualifiers
 FH Misc-difference 217..218
 FT /note="Encoded by AGTARAAGA"
 FT Misc-difference 224..225
 FT /note="Encoded by CT"
 XX
 PN WO200160849-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 16-FEB-2001; 2001WO-US05048.
 XX
 PR 16-FEB-2000; 2000US-0506286.
 XX
 PA (UYP1-) UNIV PITTSBURGH.
 XX
 PI Dowling PW, Youngner JS;
 XX
 XX WPI; 2001-522584/57.
 DR N-PSDB; AAD15724.
 XX
 PT Novel isolated equine influenza virus (wild-type and cold-adapted)
 PT proteins and viruses containing nucleic acid molecules encoding the
 PT proteins, which are useful for protecting animals from influenza virus
 PT infections -
 XX
 PS Claim 5; Page 165-166; 172pp; English.
 XX
 CC The patent discloses cold-adapted equine influenza viruses and
 CC reassortant influenza A viruses comprising at least one genome
 CC segment of such an equine influenza virus, wherein the equine
 CC influenza virus genome segment confers at least one identifying
 CC phenotype of the cold-adapted equine influenza virus, such as

CC cold adaptation, temperature sensitivity, dominant interference
 CC or attenuation. The viruses are useful for protecting animals
 CC from diseases caused by influenza viruses. They are also used
 CC as vaccines. The present sequence is equine influenza (ei) virus
 CC H3N8 Peiwt1 (wild type) NP-C-225 protein which is encoded by
 CC neiwtNP-C-679 DNA.

XX Sequence 225 AA:

Query Match 72.5%; Score 37; DB 22; Length 225;
 Best Local Similarity 66.7%; Pred. No. 35;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 DNPOHKTQL 9
 :|||:|||
 Db 39 enpahksq 47

RESULT 9
 AAW68409
 ID AAW68409 standard; Protein; 498 AA.
 XX
 AC AAW68409;
 XX
 DT 18-FEB-1999 (first entry)
 XX
 DE SIV strain H1N1 nucleoprotein.
 XX
 KW Multivalent vaccine; pig; pathogen; respiratory disease; SIV; PRRSV; HCV;
 KW digestive disease; Aujeszky's disease virus; pseudorabies virus; vaccine;
 KW swine herpesvirus 1; swine influenza virus; hog cholera virus; vector;
 KW porcine respiratory and reproductive syndrome virus; nucleoprotein;
 KW SIVSV; swine infertility and respiratory syndrome virus;
 KW Actinobacillus pleuropneumoniae.

XX Swine influenza virus.

OS FR2751224-A1.
 PN
 XX 23-JAN-1998.
 PD
 XX 19-JUL-1996; 96FR-0009338.
 PF
 XX 19-JUL-1996; 96FR-0009338.
 PR
 XX (INMR) RHONE MERIEUX SA.

XX WPI; 1998-112824/11.
 DR N-PSDB; AAV49295.

XX Multi-valent polynucleotide vaccines against porcine pathogens -
 PT consist of at least 3 plasmids able to express protective antigens
 PT from specified viruses
 XX
 PS Example 11; Fig 8; 63pp; French.

XX The invention relates to a multivalent vaccine for protecting pigs
 CC against several pathogens, especially pathogens associated with
 CC respiratory and digestive diseases. The pathogens are especially
 CC selected from Aujeszky's disease virus, swine influenza virus (SIV),
 CC porcine respiratory and reproductive syndrome virus (PRRSV), hog
 CC cholera virus (HCV) and Actinobacillus pleuropneumoniae. The vaccines
 CC are preferably composed of polynucleotide sequences encoding 3 antigens,
 CC all as part of vectors. This sequence represents the SIV strain H1N1
 CC nucleoprotein. The coding sequence was subcloned into the plasmid
 CC pVR1012 to generate plasmid pPB142 for use in the vaccine.

XX Sequence 498 AA:

Query Match 72.5%; Score 37; DB 19; Length 498;
 Best Local Similarity 66.7%; Pred. No. 81;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0; -

QY 1 DNPQHKTL 9
 Db 320 enpahksq1 328

RESULT 10
 AAW68407
 ID AAW68407 standard; Protein; 498 AA.
 AC AAW68407;
 XX
 DT 18-FEB-1999 (first entry)
 XX
 DE SIV strain H3N2 nucleoprotein.
 XX
 KW Multivalent vaccine; pig; pathogen; respiratory disease; SIV; PRRSV; HCV;
 KW digestive disease; Aujeszky's disease virus; pseudorabies virus; vaccine;
 KW porcine herpesvirus 1; swine influenza virus; hog cholera virus; vector;
 KW porcine respiratory and reproductive syndrome virus; nucleoprotein;
 KW SIRS; swine infertility and respiratory syndrome virus;
 KW Actinobacillus pleuropneumoniae.
 XX
 OS Swine influenza virus.
 XX
 PN FR2751224-A1.
 XX
 PD 23-JAN-1998.
 XX
 PF 19-JUL-1996; 96FR-0009338.
 XX
 PR 19-JUL-1996; 96FR-0009338.
 XX
 PA (INMR) RHONE MERIEUX SA.
 XX
 XX WPI; 1998-112824/11.
 DR N-PSDB; AAV49301.
 XX
 PT Multi-valent polynucleotide vaccines against porcine pathogens -
 PT consist of at least 3 plasmids able to express protective antigens
 PT from specified viruses
 XX
 PS Example 13; Fig 12; 63pp; French.
 XX
 CC The invention relates to a multivalent vaccine for protecting pigs
 CC against several pathogens, especially pathogens associated with
 CC respiratory and digestive diseases. The pathogens are especially
 CC selected from Aujeszky's disease virus, swine influenza virus (SIV),
 CC porcine respiratory and reproductive syndrome virus (PRRSV), hog
 CC cholera virus (HCV) and Actinobacillus pleuropneumoniae. The vaccines
 CC are preferably composed of polynucleotide sequences encoding 3 antigens,
 CC all as part of vectors. This sequence represents the SIV strain H3N2
 CC nucleoprotein. The coding sequence was subcloned into the plasmid
 CC pVR1012 to generate plasmid pVR132 for use in the vaccine.
 XX
 SQ Sequence 498 AA;

Query Match 72.5%; Score 37; DB 19; Length 498;
 Best Local Similarity 66.7%; Pred. No. 81;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTL 9
 Db 320 enpahksq1 328

RESULT 11
 AAW55986
 ID AAW55986 standard; Protein; 499 AA.
 AC AAW55986;

XX
 DT 24-JUL-1998 (first entry)
 XX
 DE Swinepox virus HindIII M fragment protein SEQ ID NO:231.
 XX
 KW Swinepox virus; SPV; recombinant; vaccine; immunisation; diagnosis;
 KW pseudorabies virus; feline immunodeficiency virus; FIV; heartworm;
 KW Dirofilaria immitis.
 XX
 OS Swinepox virus.
 XX
 PN WO9804684-A1.
 XX
 PD 05-FEB-1998.
 XX
 PF 25-JUL-1997; 97WO-US12212.
 XX
 PR 25-JUL-1996; 96US-0686968.
 XX
 PA (SYTR) SYNTRON CORP.
 XX
 PI Cochran MD, Junker DE;
 XX
 DR WPI; 1998-130677/12.
 DR N-PSDB; AAV26105.
 XX
 PT Recombinant swine pox virus - useful in vaccine for immunising
 PT animal against swine pox virus
 XX
 PS Disclosure; Page 409-410; 473pp; English.
 XX
 CC The present sequence represents a protein from a Swinepox virus strain
 CC Kasza isolate S-SPV-001 HindIII M fragment DNA fragment, which is used in
 CC an example from the present invention. The present invention specifically
 CC describes recombinant swinepox virus (SPV) comprising a foreign DNA (1)
 CC inserted into a SPV genome which is capable of being expressed in a host
 CC cell into which the virus is introduced, where (1) is inserted into: (a)
 CC an EcoRI site within a region corresponding to a 3.2 kb subfragment of
 CC the HindIII K fragment which contains both a HindIII and an EcoRI site,
 CC of the SPV genome, and optionally (b) an AclI site within a region
 CC corresponding to a 3.6 kb HindIII to BglII subfragment of the HindIII M
 CC fragment. The recombinant SPV can be used in a vaccine for immunising an
 CC animal against SPV. The invention also provides a method for testing a
 CC vaccine to determine whether the swine has been vaccinated with a
 CC vaccine, particularly containing S-SPV-008, or is infected with a
 CC naturally occurring wild-type pseudorabies virus. Also (1) inserted into
 CC recombinant SPV can be used in a diagnostic assay, e.g. feline
 CC immunodeficiency virus (FIV) env and gag genes and Dirofilaria immitis
 CC p39 and 22kd are useful to detect feline immunodeficiency caused by FIV
 CC and to detect heartworm caused by D. immitis respectively.
 XX
 SQ Sequence 499 AA;

Query Match 72.5%; Score 37; DB 19; Length 499;
 Best Local Similarity 66.7%; Pred. No. 81;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTL 9
 Db 320 enpahksq1 328

RESULT 12
 AAR36821
 ID AAR36821 standard; Protein; 917 AA.
 XX
 AC AAR36821;
 XX
 DT 25-AUG-1993 (first entry)
 XX
 DE PE binding/translocation domains-influenza A virus nucleoprotein.
 XX

KW Vaccine; cytotoxic T lymphocyte; CTL; Influenza A virus; NP;
 KW anti-viral agent; Pseudomonas exotoxin; fusion construct.
 XX
 OS Chimeric Pseudomonas aeruginosa.
 OS Chimeric Influenza A virus.
 XX
 FH Key Location/Qualifiers
 FT Region 2..414
 FT /note= "amino acids 2-414 of PE domains 1 and II"
 FT Region 415..912
 FT /note= "Influenza A virus Nucleoprotein"
 FT Region 913..917
 FT /note= "last 5 amino acids of PE"
 XX
 XX EP541335-A.
 PN 12-MAY-1993.
 PD
 XX 04-NOV-1992; 92EP-0310067.
 XX 08-NOV-1991; 91US-0792507.
 XX (MERI) MERCK & CO INC.
 PA
 XX Donnelly JJ, Friedman A, Howe LA, Liu MA, Marshall MS;
 PI Montgomery DL, Oliff AA, Shi X, Ulmer J;
 XX WPI; 1993-154266/19.
 DR N-PSDB; AAQ41728.
 XX
 XX Recombinant DNA encoding bacterial toxin-antigen conjugates - are
 PT useful as vaccines against viral infections, tumours and
 PT parasites
 PT
 XX Example 25; Page 63-67; 81pp; English.
 PS
 XX Plasmid pApr501 was constructed from the Influenza A virus
 CC nucleoprotein gene (NP) cloned into the EcoRI site of pBR322.
 CC A fragment containing the NP gene was obtained from the plasmid by
 CC PCR with primers that added a SacII site adjacent to the ATG codon of
 CC NP, and the last 5 amino acids of PE followed by a termination
 CC codon and an EcoRI site to the 3' end of NP. The PCR fragment was
 CC digested with SacII and EcoRI and ligated to SacII/EcoRI-digested
 CC plasmid pVC-PEM1-2 (encoding a Pseudomonas exotoxin-Influenza A
 CC virus M1 matrix protein fusion). In the resulting plasmid,
 CC pVC-PENPC5aa, the binding and translocation domains of PE are fused
 CC to the Influenza A nucleoprotein.
 XX
 SQ Sequence 917 AA:

Query Match 72.5%; Score 37; DB J4; Length 917;
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHTQL 9
 :|||:|
 -DB 734 enphkksq| 742

RESULT 13
 AAR32469
 ID AAR32469 standard; Protein; 917 AA.
 XX
 AC AAR32469;
 XX
 XX 20-JUL-1993 (first entry)
 XX
 XX PE binding and translocation domains - Influenza A nucleoprotein
 DE fusion protein.
 DE
 XX
 KW PE: Pseudomonas exotoxin; Influenza A virus; M1; matrix protein;
 KW fusion; hybrid; pVC-PENPC5aa; pApr501; pBR322; pVC-PEM1-2;

KW nucleoprotein; NP; PCR; amplification; translocation;
 KW binding; domain.
 XX
 PN EP532090-A.
 XX 17-MAR-1993.
 PD
 XX 02-SEP-1992; 92EP-0202660.
 XX 09-SEP-1991; 91US-0756249.
 XX (MERI) MERCK & CO INC.
 PA
 XX Donnelly JJ, Friedman A, Howe LA, Liu MA, Marshall MS;
 PI Montgomery DL, Oliff AA, Shi X, Ulmer J;
 XX WPI; 1993-087107/11.
 DR N-PSDB; AAQ38411.
 XX
 XX Bacterial toxin-antigen protein conjugates - to elicit cytotoxic
 PT T-lymphocyte immune response, used for preventing viral
 PT infections, e.g. by influenza virus, HIV and human
 PT papilloma virus
 PT
 XX Disclosure; Page 66-70; 85pp; English.
 PS
 XX Example 25 describes the construction of pVC-PENPC5aa.
 CC A fragment contg. the nucleoprotein (NP) of Influenza A virus was
 CC obtained from plasmid pApr501. pApr501 is the nucleoprotein gene
 CC cloned into the EcoRI site of pBR322, by PCR with oligonucleotide
 CC primers which added a SacII site adjacent to the ATG codon of NP
 CC to give the sequence of AAQ38409, and the last 5 amino acids of PE
 CC followed by a termination codon and an EcoRI site to the 3' end of
 CC NP to give the sequence shown in AAQ38410. The PCR fragment was
 CC digested with SacII and EcoRI and ligated to the plasmid pVC-PEM1-2
 CC digested with SacII and EcoRI. The resulting plasmid is named
 CC pVC-PENPC5aa. The 5' and 3' ends of the PENPC5aa insert (AAQ38411)
 CC were verified by sequencing. This construction fuses the binding
 CC and translocation domains of PE to the Influenza A nucleoprotein.
 XX
 SQ Sequence 917 AA:

Query Match 72.5%; Score 37; DB J4; Length 917;
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHTQL 9
 :|||:|
 -DB 734 enphkksq| 742

RESULT 14
 AAB15944
 ID AAB15944 standard; Protein; 104 AA.
 XX
 AC AAB15944;
 XX
 XX 05-OCT-2000 (first entry)
 XX
 XX E. coli proliferation associated protein sequence SEQ ID NO:301.
 DE
 XX Escherichia coli; E. coli; proliferation; inhibition; screening;
 KW antimicrobial; bacterial growth; antisense therapy; antibacterial.
 KW
 XX Escherichia coli.
 OS
 XX WO200044906-A2.
 PN
 XX 03-AUG-2000.
 PD
 XX 27-JAN-2000; 2000WO-US02200.
 XX

PR 27-JAN-1999; 99US-0117405.
XX (ELIT-) ELITRA PHARM INC.
XX Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;
PI Yamamoto RT, Xu HH;
XX WPI: 2000-514822/46.
DR N-PSDB; AAA65949.
XX Novel polynucleotides and polypeptides associated with microorganism
PT proliferation, used to identify inhibitors of bacterial growth and
PT proliferation, for use in antisense therapy.
XX Claim 11; Page 224; 316pp; English.
XX AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
CC sequences derived from Escherichia coli which inhibit E. coli
CC proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
CC nucleotide and protein sequences associated with E. coli proliferation.
CC AAA66056 and AAA66057 represent primers used for sequencing E. coli
CC proliferation inhibiting nucleotide inserts in an example from the
CC present invention. Methods from the present invention can be used to
CC identify a proliferation- required gene in a microorganism, by contacting
CC a microorganism with a proliferation-required gene activity inhibitory
CC nucleic acid identified in another organism, and determining if
CC inhibition occurs in the second microorganism. The nucleic acid sequences
CC identified as being required for bacterial growth and proliferation, can
CC be used for antisense therapy for killing bacteria.
XX
SQ Sequence 104 AA;

Query Match 70.6%; Score 36; DB 21; Length 104;
Best Local Similarity 85.7%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PQHKTQL 9
Db 17 pqhktel 23

RESULT 15
AAAY29160
ID AAAY29160 standard; Protein: 132 AA.
XX
AC AAAY29160;
XX
DT 25-OCT-1999 (first entry)
XX
DE Amino acid sequence of a virulence factor encoded by OXP14991.
XX Human pathogen; virulence polypeptide; virulence factor;
XW pathogenic infection; Pseudomonas aeruginosa infection.
XW
OS Pseudomonas aeruginosa.
XX
PN WO9927129-A1.
XX
PD 03-JUN-1999.
XX
PF 25-NOV-1998; 98WO-US25247.
XX
PR 25-NOV-1997; 97US-0066517.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;
PI Rahne LG, Tan M, Tsongalis J;
XX WPT: 1999-357851/30.
DR
XX Virulence factors useful in developing disease treatments
PT

XX Disclosure; Fig 3; 228pp; English.
XX
XX The present sequence represents a pseudomonas aeruginosa polypeptide
CC sequence. P. aeruginosa is an opportunistic human pathogen present in
CC soil water and plants. The specification describes virulence polypeptides
CC and nucleic acid sequence encoding such polypeptides. These sequences
CC can be used to identify a compound which is capable of decreasing the
CC expression of a pathogenic virulence factor. Compounds that inhibit
CC the expression or activity of virulence factor polypeptides can be
CC used to treat pathogenic infections, especially where the infection
CC is a P. aeruginosa infection.
CC note: the sequences given in the specification were poorly legible, and
CC in some instances assumptions were made as to the identity of the
CC residue; it is therefore possible that the sequence given below is
CC not entirely correct.
XX
SQ Sequence 132 AA;

Query Match 70.6%; Score 36; DB 20; Length 132;
Best Local Similarity 62.5%; Pred. No. 31;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DNPQHKTQ 8
Db 65 dnprhsg 72

Search completed: August 6, 2002, 16:52:55
Job time: 337 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: August 6, 2002, 16:53:56 ; Search time 51.22 seconds
(without alignments)
4.292 Million cell updates/sec

Title: US-10-020-139-2_COPY_237_245

Sequence: 1 DNPQHKTL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	72.5	498	4	US-09-232-468A-18
2	37	72.5	498	4	US-09-232-468A-24
3	37	72.5	498	4	US-08-686-968C-231
4	36	70.6	132	4	US-09-199-637A-229
5	34	66.7	412	1	US-08-102-863-11
6	34	66.7	412	5	PCT-US92-10885-11
7	33	64.7	345	2	US-08-758-621-14
8	33	64.7	345	4	US-08-107-858-14
9	32	62.7	85	4	US-08-858-207A-318
10	32	62.7	115	1	US-08-052-681-3
11	32	62.7	116	1	US-08-052-681-4
12	32	62.7	712	2	US-08-474-067-2
13	32	62.7	712	2	US-08-474-067-5
14	32	62.7	712	2	US-08-474-068A-2
15	32	62.7	712	2	US-08-474-068A-5
16	32	62.7	712	2	US-08-472-481-2
17	32	62.7	717	2	US-08-474-067-4
18	32	62.7	717	2	US-08-474-068A-4
19	32	62.7	717	2	US-08-472-481-4
20	31	60.8	734	1	US-08-276-099A-16
21	31	60.8	734	1	US-08-781-890-16
22	31	60.8	778	5	PCT-US93-03076-3
23	31	60.8	1513	5	PCT-US93-03076-2
24	30	58.8	7	3	US-08-405-647B-16
25	30	58.8	7	4	US-08-985-499-16
26	30	58.8	7	5	PCT-US96-03180-16
27	30	58.8	120	4	US-09-107-858-23

28	30	58.8	126	4	US-09-284-033-5	Sequence 5, Appli
29	30	58.8	126	4	US-08-729-834B-5	Sequence 5, Appli
30	30	58.8	169	4	US-09-300-681B-5	Sequence 5, Appli
31	30	58.8	232	4	US-09-300-681B-4	Sequence 4, Appli
32	30	58.8	323	2	US-08-747-788-2	Sequence 2, Appli
33	30	58.8	323	4	US-09-300-681B-2	Sequence 2, Appli
34	30	58.8	399	4	US-09-284-033-2	Sequence 2, Appli
35	30	58.8	399	4	US-08-729-834B-2	Sequence 2, Appli
36	30	58.8	433	4	US-09-400-208B-5	Sequence 5, Appli
37	30	58.8	620	2	US-08-419-652-7	Sequence 7, Appli
38	30	58.8	652	1	US-08-765-081-6	Sequence 6, Appli
39	30	58.8	652	3	US-09-098-082-6	Sequence 6, Appli
40	30	58.8	718	5	PCT-US95-08994-7	Sequence 7, Appli
41	30	58.8	784	4	US-09-371-913A-7	Sequence 7, Appli
42	30	58.8	794	1	US-08-393-333-2	Sequence 2, Appli
43	30	58.8	794	4	US-09-087-465-10	Sequence 10, Appli
44	30	58.8	1001	1	US-07-797-556-6	Sequence 6, Appli
45	30	58.8	1001	1	US-07-943-843-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-232-468A-18
; Sequence 18, Application US/09232468A
; Patent No. 6207165
; GENERAL INFORMATION:
; APPLICANT: AUDONNET et al.
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST PORCINE
; FILE OF INVENTION: REPRODUCTIVE AND RESPIRATORY PATHOLOGIES
; FILE REFERENCE: 454313-2230
; CURRENT APPLICATION NUMBER: US/09/232.468A
; CURRENT FILING DATE: 1999-01-05
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 498
; TYPE: PRT
; ORGANISM: swine influenza virus
US-09-232-468A-18

Query Match 72.5%; Score 37; DB 4; Length 498;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKTL 9
DB 320 ENPAHKSQ 328

RESULT 2
US-09-232-468A-24
; Sequence 24, Application US/09232468A
; Patent No. 6207165
; GENERAL INFORMATION:
; APPLICANT: AUDONNET et al.
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST PORCINE
; FILE OF INVENTION: REPRODUCTIVE AND RESPIRATORY PATHOLOGIES
; FILE REFERENCE: 454313-2230
; CURRENT APPLICATION NUMBER: US/09/232.468A
; CURRENT FILING DATE: 1999-01-05
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 498
; TYPE: PRT
; ORGANISM: swine influenza virus
US-09-232-468A-24

Query Match 72.5%; Score 37; DB 4; Length 498;

Best Local Similarity 66.7%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKQTOL 9
:|||:|:|:
Db 320 ENPAHKSOL 328

RESULT 3
US-08-686-968C-231
; Sequence 231, Application US/08686968C
; Patent No. 6221361
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; FILE REFERENCE: 39119-H/JML
; CURRENT APPLICATION NUMBER: US/08/686.968C
; CURRENT FILING DATE: 1996-07-25
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 231
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Swinepox virus
US-08-686-968C-231

Query Match 72.5%; Score 37; DB 4; Length 498;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKQTOL 9
:|||:|:|:
Db 320 ENPAHKSOL 328

RESULT 4
US-09-199-637A-229
; Sequence 229, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199.637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 229
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-229

Query Match 70.6%; Score 36; DB 4; Length 132;
Best Local Similarity 62.5%; Pred. No. 10;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNPQHKQTOL 8
:|||:|:|:
Db 65 DNARHSQ 72

RESULT 5
US-08-102-863-11
; Sequence 11, Application US/08102863
; Patent No. 5466590
; GENERAL INFORMATION:
; APPLICANT: SARIASLANI, SIMA
; TITLE OF INVENTION: CONSTITUTIVE
; TITLE OF INVENTION: EXPRESSION OF P450SOY
; TITLE OF INVENTION: AND FERREDOXIN-SOY IN
; TITLE OF INVENTION: STREPTOMYCES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; ADDRESSEE: AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0,
; SOFTWARE: Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/102.863
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/807.001
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: GALLEGS, R. THOMAS
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: CR-9000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-7342
; TELEFAX: 302-892-7949
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-102-863-11

Query Match 66.7%; Score 34; DB 1; Length 412;
Best Local Similarity 62.5%; Pred. No. 77;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHKQTOL 8
:|||:|:|:
Db 106 DDPENHTQ 113

RESULT 6
PCT-US92-10885-11
; Sequence 11, Application PC/TUS9210885
; GENERAL INFORMATION:
; APPLICANT: SARIASLANI, SIMA
; TITLE OF INVENTION: CONSTITUTIVE
; TITLE OF INVENTION: EXPRESSION OF P450SOY
; TITLE OF INVENTION: AND FERREDOXIN-SOY IN
; TITLE OF INVENTION: STREPTOMYCES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; ADDRESSEE: AND COMPANY

STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch,
MEDIUM TYPE: 1.0 MB
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10885
FILING DATE: 19921216
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: GALLECOS, R. THOMAS
REGISTRATION NUMBER: 32,692
REFERENCE/DOCKET NUMBER: CR-9000-A
TELEPHONE: 302-892-7342
TELEFAX: 302-892-7949
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 amino acids
TYPE: AMINO ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
PCT-US92-10885-11

Query Match 66.7%; Score 34; DB 5; Length 412;
Best Local Similarity 62.5%; Pred. No. 77;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DNPQHK 8
Db 106 DPEHNTQ 113

RESULT 7
US-08-758-621-14
; Sequence 14, Application US/08758621
; Patent No. 5846821
; GENERAL INFORMATION:
; APPLICANT: Guerinot, Mary Lou, and Eide, David J.
; TITLE OF INVENTION: Metal-regulated Transporters and Uses Therefor
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/758,621
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/018,578
; FILING DATE: 29-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Silveri, Jean M.
; REGISTRATION NUMBER: 39,030
; REFERENCE/DOCKET NUMBER: DCI-095CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 345 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-758-621-14

Query Match 64.7%; Score 33; DB 2; Length 345;
Best Local Similarity 83.3%; Pred. No. 97;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNPQHK 6
Db 110 DNPQHK 115

RESULT 8
US-09-107-858-14
; Sequence 14, Application US/09107858
; Patent No. 6162900
; GENERAL INFORMATION:
; APPLICANT: Guerinot, Mary Lou et al.
; TITLE OF INVENTION: METAL-REGULATED TRANSPORTERS AND USES THEREFOR
; FILE REFERENCE: DCI-099CPDV
; CURRENT APPLICATION NUMBER: US/09/107,858
; CURRENT FILING DATE: 1998-06-30
; EARLIER APPLICATION NUMBER: 08/758,621
; EARLIER FILING DATE: 1996-11-27
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-107-858-14

Query Match 64.7%; Score 33; DB 4; Length 345;
Best Local Similarity 83.3%; Pred. No. 97;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNPQHK 6
Db 110 DNPQHK 115

RESULT 9
US-08-858-207A-318
; Sequence 318, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 318:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6348328e
US-08-858-207A-318

Query Match 62.7%; Score 32; DB 4; Length 85;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DNPQH 5
DB 68 DNPQH 72
RESULT 10
US-08-052-681-3
Sequence 3, Application US/08052681
Patent No. 5314819
GENERAL INFORMATION:
APPLICANT: Kazunori YAMADA et al.
TITLE OF INVENTION: NOVEL PROTEIN HAVING NITRILE HYDRATASE
TITLE OF INVENTION: ACTIVITY AND THE GENE ENCODING THE SAME, AND A METHOD FOR PROD
TITLE OF INVENTION: FROM NITRILES VIA A TRANSFORMANT CONTAINING THE GENE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/052.681
FILING DATE: 19930427
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:

LENGTH: 115 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN: Rhizobium sp. MC12643
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-052-681-3

Query Match 62.7%; Score 32; DB 1; Length 115;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DNPQH 5
DB 77 DNPQH 81
RESULT 11
US-08-052-681-4
Sequence 4, Application US/08052681
Patent No. 5314819
GENERAL INFORMATION:
APPLICANT: Kazunori YAMADA et al.
TITLE OF INVENTION: NOVEL PROTEIN HAVING NITRILE HYDRATASE
TITLE OF INVENTION: ACTIVITY AND THE GENE ENCODING THE SAME, AND A METHOD FOR F
TITLE OF INVENTION: FROM NITRILES VIA A TRANSFORMANT CONTAINING THE GENE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM: Diskette, 5.25 inch, 500 kb
MEDIUM TYPE:

COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA: US/08/052.681
FILING DATE: 19930427
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN: Rhizobium sp. MC12643
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:

Query Match 62.7%; Score 32; DB 1; Length 116;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DNPQH 5
Db 78 DNPQH 82

RESULT 12
US-08-474-067-2
Sequence 2, Application US/08474067
Patent No. 5811518
GENERAL INFORMATION:
APPLICANT: Ranscht, Barbara
TITLE OF INVENTION: T-Cadherin Adhesion Molecule
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,067
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/213,361
FILING DATE: 14-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/607,293
FILING DATE: 30-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1682
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-474-067-2

Query Match 62.7%; Score 32; DB 2; Length 712;
Best Local Similarity 100.0%; Pred. No. 3,1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 QHKTQL 9
Db 2 QHKTQL 7

RESULT 13
US-08-474-067-5
Sequence 5, Application US/08474067
Patent No. 5811518
GENERAL INFORMATION:
APPLICANT: Ranscht, Barbara
TITLE OF INVENTION: T-Cadherin Adhesion Molecule
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,067
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/213,361
FILING DATE: 14-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/607,293
FILING DATE: 30-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1683
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-474-067-5

Query Match 62.7%; Score 32; DB 2; Length 712;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 QHKTOL 9
DB 2 QHKTOL 7

RESULT 14
US-08-474-068A-2
Sequence 2, Application US/08474068A
Patent No. 5837525
GENERAL INFORMATION:
APPLICANT: Ranscht, Barbara
TITLE OF INVENTION: T-Cadherin Adhesion Molecule
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,068A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/213,361
FILING DATE: 14-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/607,293
FILING DATE: 30-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1683
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-474-068A-5

TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-474-068A-2

Query Match 62.7%; Score 32; DB 2; Length 712;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 QHKTOL 9
DB 2 QHKTOL 7

RESULT 15
US-08-474-068A-5
Sequence 5, Application US/08474068A
Patent No. 5837525
GENERAL INFORMATION:
APPLICANT: Ranscht, Barbara
TITLE OF INVENTION: T-Cadherin Adhesion Molecule
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,068A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/213,361
FILING DATE: 14-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/607,293
FILING DATE: 30-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1683
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-474-068A-5

Query Match 62.7%; Score 32; DB 2; Length 712;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 QHKTOL 9
DB 2 QHKTOL 7

Search completed: August 6, 2002, 16:53:57
Job time: 289 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 16:52:58 ; Search time 32.88 Seconds
(without alignments)
293.223 Million cell updates/sec

Title: US-10-020-139-2

Perfect score: 1233

Sequence: 1 MLQLKLVLLCGVLTGTSSES.....NVLIQVVDNPOKHTQLOTLLI 249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	364.5	29.6	235	1 PSP_MOUSE	P07743 mus musculus
2	194.5	15.8	278	1 PLUN_MOUSE	P97361 mus musculus
3	183.5	14.9	256	1 PLUN_HUMAN	Q9np53 homo sapien
4	109	8.8	1769	1 YK9_YEAST	P42945 saccharomyc
5	106	8.6	295	1 ALF_STACA	Q07159 staphylococ
6	104.5	8.5	1727	1 ALM1_SCHPO	Q9utk5 schizosacch
7	104	8.4	1616	1 P200_MYCGE	Q49429 mycoplasma
8	104	8.4	2710	1 TOXA_CLODI	P16154 clostridium
9	100	8.1	1057	1 EG5_HUMAN	P52732 homo sapien
10	99	8.0	679	1 DNLJ_HAFIN	P43813 haemophilus
11	98.5	8.0	1616	1 RRPO_TOMK2	P89676 tomato mosa
12	98.5	8.0	1616	1 RRPO_TOM1	Pu3587 tomato mosa
13	98.5	8.0	1616	1 RRPO_TOM1	Q9yxd6 tomato mosa
14	97	7.9	757	1 DNMI_YEAST	P54861 saccharomyc
15	97	7.9	958	1 YG7_YEAST	P53076 saccharomyc
16	96	7.8	1005	1 RA50_METJA	Q58718 methanococ
17	95	7.7	1531	1 YQ38_CAEEL	Q09459 caenorhabdi
18	94.5	7.7	1729	1 RRP5_YEAST	Q05022 saccharomyc
19	94	7.6	462	1 NIFK_METMP	P71527 methanococ
20	94	7.6	481	1 LBP_HUMAN	P18428 homo sapien
21	94	7.6	490	1 ILVC_BUCAI	P57655 buchnera ap
22	93.5	7.6	868	1 N180_YEAST	P33420 saccharomyc
23	93.5	7.6	1038	1 YK3_YEAST	P36097 saccharomyc
24	92.5	7.5	529	1 VGLF_SV5	P04849 simian viru
25	92.5	7.5	668	1 RRPO_BACSU	P42571 bacillus su
26	92.5	7.5	1616	1 HAP1_HAFIN	Q9qit8 tomato mosa
27	92	7.5	1409	1 HAP1_HAFIN	P44596 haemophilus
28	92	7.5	1957	1 YS86_SCHPO	Q10411 schizosacch
29	92	7.5	2329	1 YS86_CAEEL	Q09624 caenorhabdi
30	91.5	7.4	338	1 RLA0_METTL	O52705 methanococ
31	91.5	7.4	4967	1 RYR2_HUMAN	Q92736 homo sapien
32	91	7.4	368	1 ISPG_LISMO	P58668 listeria mo
33	91	7.4	2376	1 YIM5_YEAST	P40468 saccharomyc

ALIGNMENTS

RESULT 1

ID	PSP_MOUSE	STANDARD	PRT	235 AA
AC	P07743			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-AUG-1988 (Rel. 08, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Parotid secretory protein precursor (PSP).			
GN	PSP.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Parotid gland;			
RX	MEDLINE=85215456; PubMed=2582349;			
RA	Madsen H.O., Hjorth J.P.;			
RT	"Molecular cloning of mouse PSP mRNA."			
RL	Nucleic Acids Res. 13:1-13(1985).			
RN	[2]			
RP	SEQUENCE OF 1-87 FROM N.A.			
RC	STRAIN=C3H; TISSUE=Spleen;			
RX	MEDLINE=87004556; PubMed=2428613;			
RA	Poulsen K., Jakobsen B.K., Mikkelsen B.M.,			
RT	Nielsen J.T., Hjorth J.P.;			
RL	"Coordination of murine parotid secretory protein and salivary amylase expression."			
RL	EMBO J. 5:1891-1896(1986).			
CC	FUNCTION: PSP IS THE MOST ABUNDANT PROTEIN IN THE PAROTID GLAND.			
CC	ITS FUNCTION IS NOT KNOWN; HOWEVER, ITS PRODUCTION IS COORDINATED WITH THAT OF SALIVARY AMYLASE.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X01697; CAA25846.1;			
DR	EMBL; M26807; AAA40009.1;			
DR	EMBL; M26806; AAA40009.1; JOINED.			
DR	PIR; A23031; SQMS.			
DR	MGI; 97787; PSP.			
DR	Parotid gland; Signal.			
FT	SIGNAL 1 20			
FT	CHAIN 21 235			
SQ	SEQUENCE 235 AA; 24753 MW; 23311BAE1E6E2EF3 CRC64;			

Query Match 29.6%; Score 364.5; DB 1; Length 235;
Best Local Similarity 33.8%; Pred. No. 2.1e-20;
Matches 81; Conservative 62; Mismatches 80; Indels 17; Gaps 3;

QY 1 MLQWKLVLCGLVLTGTSTESLDNLGNSVVDKLEPVLHEGLETVDTNLTGKILEKLV 60
 FT CHAIN 20 278
 FT DOMAIN 23 52
 FT REPEAT 23 28
 FT REPEAT 30 36
 FT REPEAT 39 44
 FT REPEAT 47 52
 FT CARBOHYD 182 182
 FT CARBOHYD 228 228
 SQ SEQUENCE 278 AA; 28611 MW; 7F40BED959188FB CHK64;

Query Match 15.8%; Score 194.5; Db 1; Length 278;
 Best Local Similarity 23.5%; Pred. No. 1.2e-07;
 Matches 68; Conservative 49; Mismatches 111; Indels 61; Gaps 9;

RESULT 2
 PLUN_MOUSE STANDARD; PRT; 278 AA.
 ID PLUN_MOUSE
 AC P97361;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Protein Plunc precursor (Palate lung and nasal epithelium clone protein).
 DE protein.
 GN Plunc.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=21290678; PubMed=11969672;
 RA LeClair E.E., Nguyen L., Bingle L., MacGowan A., Singleton V.,
 Ward S.J., Bingle C.D.;
 RA "Genomic organization of the mouse plunc gene and expression in the
 developing airways and thymus";
 RT Biochem. Biophys. Res. Commun. 284:792-797(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR; TISSUE=Palate;
 RX MEDLINE=9240770; PubMed=10224143;
 RA Weston W.M., LeClair E.E., Trzyna W., McHugh K.M., Nugent P.,
 Laferty C.M., Ma L., Tuan R.S., Greene R.M.;
 RA "Differential display identification of plunc, a novel gene expressed
 in embryonic palate, nasal epithelium, and adult lung";
 RL J. Biol. Chem. 274:13698-13703(1999).
 RN [3]
 RP ERRATUM.
 RA Weston W.M., LeClair E.E., Trzyna W., McHugh K.M., Nugent P.,
 Laferty C.M., Ma L., Tuan R.S., Greene R.M.;
 RL J. Biol. Chem. 275:8262-8262(2000).
 CC -!- FUNCTION: May be involved in the airway inflammatory response
 after exposure to irritants. May be associated with tumor
 progression (By similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -!- TISSUE SPECIFICITY: Upper airways, nasopharyngeal regions and
 thymus.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
 or send an email to license@isb-sib.ch).

EMBL: AF356785; AAK63069.1;
 DR EMBL: U69172; AAB63256.1;
 DR MGD: MGI:1338036; Plunc.
 XW Signal; Repeat.

FT SIGNAL 1 19
 FT CHAIN 20 278
 FT DOMAIN 23 52
 FT REPEAT 23 28
 FT REPEAT 30 36
 FT REPEAT 39 44
 FT REPEAT 47 52
 FT CARBOHYD 182 182
 FT CARBOHYD 228 228
 SQ SEQUENCE 278 AA; 28611 MW; 7F40BED959188FB CHK64;

Query Match 15.8%; Score 194.5; Db 1; Length 278;
 Best Local Similarity 23.5%; Pred. No. 1.2e-07;
 Matches 68; Conservative 49; Mismatches 111; Indels 61; Gaps 9;

QY 1 MLQWKLVLCGLVLTGTSTESLDNLGNSVVDKLEPVLHEGLETVDTNLTGKILEKLV 60
 Db 1 MFLVGSILVLCGLLAHSTAQAGLPPLGOGPPLPLNGQPLPLNGUQLPLAOGPLIAY 60
 QY 30 -----SNVVD-----KLEPVLHEGLETVDTNLTGKILEKLVKLVGLOKSSAMUAKAKAE 80
 Db 61 SPALPSNPTDLLAGKFTDALSGGL-----LSGGL-----LGILENIPLLDVIKSGGNN 108
 QY 81 AEKLLNVISKL---LPTNTDIFGLKITSNLSILDVKAEPIDGCKGLNLSFPVTANTVAG 137
 Db 109 SNGVGLGGLKLTSSVPLNNILDKITDPQLLEGLGVSPDGRRLYVTPICGLTLNVNM 168
 QY 138 PIQGIINLKASDLLPAVITETDPQ--THQPVAVIGECASDPTSISLLOKHSHIINK 195
 Db 169 PVGSGLLQAVKLNTAEVLAVKQNGRIH---LVLDGCTHSPGSLKISLLNGVTP-VQS 224
 QY 196 FVNSVINTLSTVSSLLQKEICPLIRIFIHSLDNNVIOQVVDNPOHKTO 244
 Db 225 FVDNLGTGLTKVPELTQGVKCPVNLVSLGSLDVTLVHNTAELLHGLQ 273

RESULT 3
 PLUN_HUMAN STANDARD; PRT; 256 AA.
 ID PLUN_HUMAN
 AC Q9NP55; Q9NZT0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Protein Plunc precursor (Palate lung and nasal epithelium clone protein) (Lung specific X protein) (Nasopharyngeal carcinoma-related protein) (Tracheal epithelium enriched protein).
 DE PLUNC OR LUNX.
 GN Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20472055; PubMed=11018263;
 RA Bingle C.D., Bingle L.;
 RT "Characterization of the human plunc gene, a gene product with an
 upper airways and nasopharyngeal restricted expression pattern";
 RL Biochim. Biophys. Acta 1493:363-367(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=21150203; PubMed=11251963;
 RA Iwao K., Watanabe T., Fujiwara Y., Takami K., Kodama K.,
 Higashiyama M., Yokouchi H., Ozaki K., Monden M., Tanigami A.;
 RT "Isolation of a novel human lung-specific gene, LUNX, a potential
 molecular marker for detection of micrometastasis in non-small-cell
 lung cancer";
 RL Int. J. Cancer 91:433-437(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA He Z.W., Yao K.T., Xu L.G., Ren C.P., Lan K., Xie L., Zhang L.,
 Liu W.D., Zhou W., Wang L.;


```

QY 80 BAELNNV1--SKLLPTNTDI-----FGLKISNSIILDKAEPIDDKGKGLNLSPPVTAN 132
Db 379 EVRLITDILVLESEILDKSOLVLEFEYFISINEDIVLKCL-----KSLGLT----- 425
QY 133 VTWAGPIIQQIINLKASLDLIJAVTITETD--POTHOPVAVLGEACASDPTSISLSLDDKHS 190
Db 426 -----GELFEIRLTSLFTNADYNTDIVKOLSPV-----ETTKKDTASTQTFEDKHS 473
QY 191 QIIN-----KFVNSVINTLKSTVSSLLQKEICP 218
Db 474 ELINTNVSMITFEGYKRYKVLSTFAIGCKGYKASSEFUTSPFTTLESRTITFLLRVTISP 533
QY 219 LIRIFHSLOVNVIOQVVDNPNQHKTLQTLI 249
Db 534 AAPTALKILSLNNIAKYINSIEKEVNIFTLV 564

RESULT 5
ALF_STACA STANDARD; PRT; 295 AA.
AC Q07159;
DT 01-OCT-1994 (rel. 30, Created)
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE Fructose-bisphosphate aldolase class I (EC 4.1.2.13) (FEP aldolase).
GN FDA.
OS Staphylococcus carnosus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1281;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TM300;
RX MEDLINE=94042930; PubMed-8226699;
RA Witke C., Goetz F.;
RT "Cloning, sequencing, and characterization of the gene encoding the
RT class I fructose-1,6-bisphosphate aldolase of Staphylococcus
RT carnosus";
RL J. Bacteriol. 175:7495-7499(1993).
RN [2]
RP SEQUENCE OF 1-7.
RX MEDLINE=94042930; PubMed-8226699;
RA Kula M.R., Brockamp H.P.;
RL Unpublished results, cited by:
RL Witke C., Goetz F.;
RL J. Bacteriol. 175:7495-7499(1993).
CC -|- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate -> glyceralone
CC phosphate + D-glyceraldehyde 3-phosphate.
CC -|- PATHWAY: SIXTH STEP IN GLYCOLYSIS.
CC -|- SIMILARITY: BELONGS TO CLASS I FRUCTOSE-BISPHOSPHATE ALDOLASE
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X71729; CAA50663.1; .
CC PIR; S33358; S33358.
CC PIR; A49943; A49943.
CC HSP; P14223; IA5C.
CC InterPro; IPR000741; Aldolase_I.
CC Pfam; PF00274; glycolytic_enzy; 1.
CC ProDom; PD001128; Aldolase-I; 1.
CC PROSITE; PS00159; ALDOLASE_CLASS_I; FALSE_NEG.
CC Lysase; Schiff base; Glycolysis.
CC INET 0
CC BINDING 211 211 SCHIFF-BASE WITH DIHYDROXYACETONE-P
CC FT BINDING 295 AA; 32720 MW; A189E75574F1FCC0 CRC64;
CC SEQUENCE 295 AA; 32720 MW; A189E75574F1FCC0 CRC64;
CC SQ

```

```

Query Match 8.6%; Score 106; DH 1; Length 245;
Best Local Similarity 25.9%; Pred. NO. 0.5; Mismatches 64; Indels 48; Gaps 9;
Matches 50; Conservative 31;

QY 52 KGILEKLVOLGVLOKSSAWOLAKQAQAEKILNNVISKLLPTNTDIFGLKISNSIILD 111
Db 93 KGIVPFLKVDKGLAEADGVOLMK-PIPDLDKLLDKA-----NERGIFGTRK-RSNILE 144
QY 112 VKAEPIDGKGLNLSFPVTANTVAVG--PIIGQIIN-----IKASLD-- 151
Db 145 NNKEAIE--KVVKQQFEVAKETIAAGLVPIIEPEVNNAKDKKAEIANIAEAKAKLNL 202
QY 152 -----LITAVTITET-----DPOTHOPVAVLGEACASDPTSISLSLDDKHSQIIN 196
Db 203 KQDQVVMILKLTPTKVNAYSELIEHPQVIRVVALSGCYSHDEAN--KILKQKIGLIASF 259
QY 197 VNSVINTLKSTVS 209
Db 260 SRALVSDLNAAQOS 272

RESULT 6
ALM1_SCHPO STANDARD; PRT; 1727 AA.
AC Q9UTK5; Q13313; Q9UTT8;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Abnormal long morphology protein 1 (Sp8).
GN ALM1 OR SPAC1486.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA McDougall R.C., Rajandream M.A., Barrell B.G., Seeger K., Harris D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.
RC STRAIN=972;
RX MEDLINE=20123449; PubMed=10660053;
RA Jimenez M., Petit F., Gancedo C., Goday C.;
RT "The alm1+ gene from Schizosaccharomyces pombe encodes a coiled-coil
RT protein that associates with the medial region during mitosis.";
RL Mol. Gen. Genet. 262:921-930(2000).
RN [3]
RP SEQUENCE OF 644-834 FROM N.A.
RC STRAIN=968 H90;
RA Ding D.;
RT "Generation and analysis of GFP-gene fusion library of fission
RT yeast.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,
CC CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN
CC CYTOKINESIS.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AL133357; CAB62414.1; .
CC EMBL; AF010473; AAB65416.1; ALT_INIT.
CC EMBL; AB028012; BAA87316.1; .
CC DR
CC DR

```


Nucleic Acids Res. 18:1629-1630(1990).

[2] SEQUENCE FROM N.A.

RA STRAIN=VPI 10463;

RA MEDLINE=90129305; PubMed=2105276;

RA Dove C.H., Wang S.Z., Price S.B., Phelps C.J., Iyerly D.M.,

RA Wilkins T.W., Johnson J.L.;

RA "Molecular characterization of the Clostridium difficile toxin A

gene";

RA Infect. Immun. 58:480-488(1990).

[3] SEQUENCE FROM N.A.

RA STRAIN=VPI 10463;

RA von Eichel-Streiber C.;

RA Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.

CC -!- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA

CC REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE

CC DIFFERENT OLIGOPEPTIDES.

CC -!- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN

CC ENTEROTOXIN CALLED A AND CYTOTOXIN B. ONLY AFTER THE ENTERAL

CC DELIVERY OF THE ENTEROTOXIN A MAY THE CHARACTERISTIC DISEASE

CC CALLED PSEUDOMEMBRANOUS COLITIS BE INDUCED.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X51797; CAA35094.1; -

DR ENBL; M30307; AAZ3283.1; -

DR EMBL; X32982; CAA63504.1; -

DR PIR; S08638; S08638.

DR InterPro: IPR002479; CW_binding.

DR Pfam; PF01473; CW_binding_1; 31.

KW Toxin; Enterotoxin.

SQ SEQUENCE 2710 AA; 308052 MW; 0A6E52CE84C14421 CRC64;

Query Match 8.4%; Score 104; DB 1; Length 2710;

Best Local Similarity 20.3%; Pred. No. 12;

Matches 56; Conservative 63; Mismatches 109; Indels 48; Gaps 11;

QY 6 KLVLCGVLTGTSELDNLGNLSNVV-----DKLEPLVHEGLTVDNTLK--GILEK 57

DB 807 KTLLEDASVSPDKFNLNLKLNIESSIGDYIYEKLEPKVNIHNSIDDLDEFNLLEN 866

QY 58 LKVDGLVQLK-----SSAQLAKQAQAEKLLN--NVISKLLPTNTDIFG----- 101

DB 867 VSDLEYEKLNLDKYLISFEDTSKNSTYSVRINKSGESVYVTEKEIFSKYSEH 926

QY 102 -----LKISNLSLDVKAEPIDBCKGNLSFPVTANVTAVGPIIIGQINLKASLDLTLAV 156

DB 927 ITRKISTIKNSLITDVNGNLNLDN---IQLDHTSQVNTLNAFFIQSLIDYSSNKVDNL 983

QY 157 TIEPTDPTHPVAVLGCASDPTISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEI 216

DB 984 STSVKVVQLV---AQLESTGLNTIYLIQVLN-----LISNAVNDTINVLPTITEGI----- 1031

QY 217 CPLRIIFSHLDVN-VTQOVVD--NPQHTQLQTLI 249

DB 1032 -PIVSTILDGINIGAAIKELLDHEDPLKKLEAKV 1066

RESULT 9

ID_EGS_HUMAN

AC P52332; O15716; STANDARD; PRT; 1057 AA.

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Kinesin-related motor protein Eg5 (Kinesin-like spindle protein HKSP)

DE (Thyroid receptor interacting protein 5) (TRIP5) (Kinesin-like protein

DE 1).

DE GN KNSL1 OR EG5.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID:9606;

RN [1]

RP SEQUENCE FROM N.A., PHOSPHORYLATION SITE THR-927, AND MUTAGENESIS.

RA MEDLINE=96128120; PubMed=8548803;

RA Blangy A., Lane H.A., D'Herin P., Harper M., Kress M., Nigg E.A.;

RA "Phosphorylation by p34cdc2 regulates spindle association of human

RA Eg5, a kinesin-related motor essential for bipolar spindle formation

RA in vivo.";

RL Cell 83:1159-1169(1995).

RN [2]

RP SEQUENCE FROM N.A.

RA MEDLINE=98369052; PubMed=9701554;

RA Whitehead C.M., Rattner J.B.;

RA "Expanding the role of HsEg5 within the mitotic and post-mitotic

RA phases of the cell cycle.";

RL J. Cell Sci. 111:2551-2561(1998).

RN [3]

RP SEQUENCE OF 819-868 FROM N.A.

RA MEDLINE=95295737; PubMed=7776974;

RA Lee J.W., Choi H.-S., Gyuris J., Brent R., Moore D.D.;

RA "Two classes of proteins dependent on either the presence or absence

RA of thyroid hormone for interaction with the thyroid hormone

RA receptor.";

RL Mol. Endocrinol. 9:243-254(1995).

CC -!- FUNCTION: MOTOR PROTEIN REQUIRED FOR ESTABLISHING A BIPOLAR

CC SPINDLE. BLOCKING OF EG5 PREVENTS CENTROSOME MIGRATION AND ARREST

CC CELLS IN MITOSIS WITH MONOASTRAL MICROTUBULE ARRAYS.

CC -!- SUBUNIT: INTERACTS WITH THE THYROID HORMONE RECEPTOR IN THE

CC PRESENCE OF THYROID HORMONE.

CC -!- PTM: PHOSPHORYLATED EXCLUSIVELY ON SERINE DURING S PHASE, BUT ON

CC BOTH SERINE AND THR-927 DURING MITOSIS, SO CONTROLLING THE

CC ASSOCIATION OF EG5 WITH THE SPINDLE APPARATUS (PROBABLY DURING

CC EARLY PROPHASE).

CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BMC

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X85137; CAA59449.1; -

DR EMBL; U37426; AAA86132.1; -

DR EMBL; L40372; AAC41739.1; -

DR HSSP; P33176; 1BG2.

DR MIM; 148760; -

DR InterPro: IPR001752; kinesin.

DR Pfam; PF00225; kinesin; 1.

DR PRINTS; PR00380; KINESINHEAVY.

DR SMART; SM00129; KISC; 1.

DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.

DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.

KW Motor protein; Microtubules; ATP-binding; Coiled coil; Mitosis;

KW Phosphorylation.

FT DOMAIN 16 363 KINESIN-MOTOR (BY SIMILARITY).

FT DOMAIN 364 480 COILED COIL (POTENTIAL).

FT DOMAIN 737 764 COILED COIL (POTENTIAL).

FT NP_BIND 105 112 ATP (POTENTIAL).

FT MOD_RES 927 927 PHOSPHORYLATION (BY CDC2).

FT MUTAGEN 927 927 T->A: NO MITOTIC PHOSPHORYLATION. NO

FT BINDING TO SPINDLE APPARATUS.

FT CONFLICT 674 676 RNS -> EL (IN REF. 2).

FT SEQUENCE 1057 AA; 119273 MW; E322F2141BEF1601 CRC64;

```

CC CC
CC or send an email to license@isb-sib.ch).
DR EMBL; U32789; AAC22753.1; -
DR HSSP; O87703; 1B04.
DR TIGR; H11100; -
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR001679; DNA_ligase_N.
DR InterPro; IPR004150; DNA_ligase_OB.
DR InterPro; IPR004149; DNA_ligase_ZBD.
DR InterPro; IPR004445; HHH.
DR InterPro; IPR003583; HHH_1.
DR Pfam; PF00533; BRCT; 1.
DR Pfam; PF01653; DNA_ligase_N; 1.
DR Pfam; PF03120; DNA_ligase_OB; 1.
DR Pfam; PF03119; DNA_ligase_ZBD; 1.
DR Pfam; PF06633; HHH; 1.
DR ProDom; P003944; DNA_ligase_N; 1.
DR SMART; SM00292; BRCT; 1.
DR SMART; SM00278; HHH; 2.
DR SMART; SM00532; LIGANC; 1.
DR PROSITE; PS01072; BRCT; 1.
DR PROSITE; PS01055; DNA_LIGASE_N1; 1.
DR PROSITE; PS01056; DNA_LIGASE_N2; 1.
KW Ligase; DNA repair; DNA replication; NAD; Complete proteome.
FT DOMAIN 601 679 BRCT.
FT BINDING 125 125 AMP (BY SIMILARITY).
SQ SEQUENCE 679 AA; 75188 MW; 4C0C23E25D70FF52 CRC64;

Query Match 8.0%; Score 99; DB 1; Length 679;
Best Local Similarity 23.1%; Pred. No. 4.9; Mismatches 71; Indels 68; Gaps 11;
Matches 54; Conservative 41;

QY 24 NLGNDLSNVVDKLEPVHLEGLETVDTNLTGILEKLVGLQKSSAMOLA-KOKAQAE 82
DB 289 SLGYDIDGTGVLKINDI-----ALQNELGFTSKAPRWAIYAKFPAQBEL 331

QY 83 KLNNV-----ISKLLPTNTDFGLKISNLSLDVKAEPIDGKGLNLSFPVT 130
DB 332 TLNDYEFQVGRGTGATPVAKLEP-VFVAGVTVSNAYLHNG-----DEIERLNIAGDT 384

QY 131 ANVTAGPIIGIILNKASLDLLTAVTITETDPTQHPVAVLGECASTDPTSLSLLDKHS 190
DB 385 VVIRAGDVIPQII-----GVLHERREDNAKPIIF-----PTNCPVC-----DS 423

QY 191 QIINKFVSVNTLSTVSSLLQKEICPLIRIFI--HSLDVN-----VIOQVWD 237
DB 424 QIIRIEGEAVARTCTGLFCAQORKE---ALKHFVSKKAMDIDGVGKLEIOLVD 474

RESULT 11
RRPO_TUMK2 STANDARD; PRT: 1616 AA.
ID RRPO_TUMK2
AC P89676; P90349;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE RNA-directed RNA polymerase (EC 2.7.7.48) (183 kDa protein) [Contains:
DE Methyltransferase/RNA helicase (MT/HEL) (126 kDa protein)]
OS Tobacco mosaic virus (strain Kazakh K2) (TMV) (TMV strain K2).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=138312;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98012318; PubMed=9454068;
RA Belenovich E.V., Genetsov E.V., Novikov V.K., Zavrjev S.K.;
RT Properties and structure of the tobacco mosaic virus strain K2
RL genome.
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.

```

```

CC CC
CC or send an email to license@isb-sib.ch).
DR EMBL; U32789; AAC22753.1; -
DR HSSP; O87703; 1B04.
DR TIGR; H11100; -
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR001679; DNA_ligase_N.
DR InterPro; IPR004150; DNA_ligase_OB.
DR InterPro; IPR004149; DNA_ligase_ZBD.
DR InterPro; IPR004445; HHH.
DR InterPro; IPR003583; HHH_1.
DR Pfam; PF00533; BRCT; 1.
DR Pfam; PF01653; DNA_ligase_N; 1.
DR Pfam; PF03120; DNA_ligase_OB; 1.
DR Pfam; PF03119; DNA_ligase_ZBD; 1.
DR Pfam; PF06633; HHH; 1.
DR ProDom; P003944; DNA_ligase_N; 1.
DR SMART; SM00292; BRCT; 1.
DR SMART; SM00278; HHH; 2.
DR SMART; SM00532; LIGANC; 1.
DR PROSITE; PS01072; BRCT; 1.
DR PROSITE; PS01055; DNA_LIGASE_N1; 1.
DR PROSITE; PS01056; DNA_LIGASE_N2; 1.
KW Ligase; DNA repair; DNA replication; NAD; Complete proteome.
FT DOMAIN 601 679 BRCT.
FT BINDING 125 125 AMP (BY SIMILARITY).
SQ SEQUENCE 679 AA; 75188 MW; 4C0C23E25D70FF52 CRC64;

Query Match 8.0%; Score 99; DB 1; Length 679;
Best Local Similarity 23.1%; Pred. No. 4.9; Mismatches 71; Indels 68; Gaps 11;
Matches 54; Conservative 41;

QY 24 NLGNDLSNVVDKLEPVHLEGLETVDTNLTGILEKLVGLQKSSAMOLA-KOKAQAE 82
DB 289 SLGYDIDGTGVLKINDI-----ALQNELGFTSKAPRWAIYAKFPAQBEL 331

QY 83 KLNNV-----ISKLLPTNTDFGLKISNLSLDVKAEPIDGKGLNLSFPVT 130
DB 332 TLNDYEFQVGRGTGATPVAKLEP-VFVAGVTVSNAYLHNG-----DEIERLNIAGDT 384

QY 131 ANVTAGPIIGIILNKASLDLLTAVTITETDPTQHPVAVLGECASTDPTSLSLLDKHS 190
DB 385 VVIRAGDVIPQII-----GVLHERREDNAKPIIF-----PTNCPVC-----DS 423

QY 191 QIINKFVSVNTLSTVSSLLQKEICPLIRIFI--HSLDVN-----VIOQVWD 237
DB 424 QIIRIEGEAVARTCTGLFCAQORKE---ALKHFVSKKAMDIDGVGKLEIOLVD 474

RESULT 11
RRPO_TUMK2 STANDARD; PRT: 1616 AA.
ID RRPO_TUMK2
AC P89676; P90349;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE RNA-directed RNA polymerase (EC 2.7.7.48) (183 kDa protein) [Contains:
DE Methyltransferase/RNA helicase (MT/HEL) (126 kDa protein)]
OS Tobacco mosaic virus (strain Kazakh K2) (TMV) (TMV strain K2).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=138312;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98012318; PubMed=9454068;
RA Belenovich E.V., Genetsov E.V., Novikov V.K., Zavrjev S.K.;
RT Properties and structure of the tobacco mosaic virus strain K2
RL genome.
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.

```

CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC

DR EMBL: Z92909; CAB07439.1; -;
DR EMBL: Z92909; CAB07438.1; -;
DR InterPro: IPR001788; RNA_dep_RNApol2.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR000606; Viral_helicase1.
DR Pfam: PF00978; RNA_dep_RNApol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 833 840 ATP (POTENTIAL).
SQ SEQUENCE 1616 AA; 183614 MW; C2F5CE0C8C965336 CRC64;

Query Match 8.0%; Score 98.5; DB 1; Length 1616;
Best Local Similarity 20.4%; Pred. No. 16;
Matches 40; Conservative 44; Mismatches 79; Indels 33; Gaps 7;
QY 57 KLVKVDGLGVQLKSSAWQLAKQAQAEKLLNNVSKLLPTNTDIFGLKISNLSLIDVKAEP 116
DB 536 KMSVDMPLVD-----IRKKMEETEEMYNALSLSVLKNSDKFDVDFSQMCSQSLVDP 588
QY 117 IDDGK-----GLNLSF--PVTANVTAGPIGOINLKASDLLTAVTIETDPQT 164
DB 589 MTAARVIVAVASNSGTLTPEQTEANVALA-----LQDSEKASDGLVVTSRDV 639
QY 165 HOPVAVLGECAADPTIS--LSLLDKHSQIINKFVNSVINTLSTVSSLLQKEICPLIRI 222
DB 640 EEP-SIKSMARGELQIAGLSDVPESSTYRSEIESLEQFHMTATSSLHKOMCSI--V 696
QY 223 FIHSLDVNVVIOOVVDN 238
DB 697 YTGPLKVVQMKNFIDS 712

RESULT 12
RPO_TOML
ID RPO_TOML STANDARD; PRT: 1616 AA.
AC P03587; Q41352;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RNA-directed RNA polymerase (EC 2.7.7.48) (183 kDa protein) [Contains:
DE Methyltransferase/RNA helicase (MT/HEL) (126 kDa protein)].
OS Tomato mosaic virus (strain 1) (TOMV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=12252;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85157522; PubMed=6549393;
RA Ohno T., Aoyagi M., Yamashita Y., Saito H., Ikawa S., Meshi T.,
RA Okada Y.
RT "Nucleotide sequence of the tobacco mosaic virus (tomato strain)
RT genome and comparison with the common strain genome."
RL J. Biochem. 96:1915-1923(1984).
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC

DR EMBL: X02144; CAA26085.1; -;
DR EMBL: X02144; CAA26082.1; -;
DR PIR: A04195; WMTM81.
DR InterPro: IPR001788; RNA_dep_RNApol2.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR000606; Viral_helicase1.
DR Pfam: PF00978; RNA_dep_RNApol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 833 840 ATP (POTENTIAL).
SQ SEQUENCE 1616 AA; 183564 MW; A8EC8929B5CF7CAF CRC64;

Query Match 8.0%; Score 98.5; DB 1; Length 1616;
Best Local Similarity 20.4%; Pred. No. 16;
Matches 40; Conservative 44; Mismatches 79; Indels 33; Gaps 7;
QY 57 KLVKVDGLGVQLKSSAWQLAKQAQAEKLLNNVSKLLPTNTDIFGLKISNLSLIDVKAEP 116
DB 536 KMSVDMPLVD-----IRKKMEETEEMYNALSLSVLKNSDKFDVDFSQMCSQSLVDP 588
QY 117 IDDGK-----GLNLSF--PVTANVTAGPIGOINLKASDLLTAVTIETDPQT 164
DB 589 MTAARVIVAVASNSGTLTPEQTEANVALA-----LQDSEKASDGLVVTSRDV 639
QY 165 HOPVAVLGECAADPTIS--LSLLDKHSQIINKFVNSVINTLSTVSSLLQKEICPLIRI 222
DB 640 EEP-SIKSMARGELQIAGLSDVPESSTYRSEIESLEQFHMTATSSLHKOMCSI--V 696
QY 223 FIHSLDVNVVIOOVVDN 238
DB 697 YTGPLKVVQMKNFIDS 712

RESULT 13
RPO_TOML
ID RPO_TOML STANDARD; PRT: 1616 AA.
AC Q9YK06; Q9WJ37;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RNA-directed RNA polymerase (EC 2.7.7.48) (183 kDa protein) [Contains:
DE Methyltransferase/RNA helicase (MT/HEL) (126 kDa protein)].
OS Tomato mosaic virus (strain S-1) (TOMV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=138314;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhou X., Xue C., Chen Q., Qi Y., Li D.;
RT "Complete nucleotide sequence of a Chinese isolate of tomato mosaic
RT virus."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC Un* European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AJ132845; CAB36997.1; -
 DR EMBL: AJ132845; CAB36998.1; -
 DR InterPro: IPR001788; RNA_dep_RNApol2.
 DR InterPro: IPR002588; V_methyltransferase.
 DR InterPro: IPR000606; Viral_helicase1.
 DR Pfam: PF00978; RNA_dep_RNApol2; 1.
 DR Pfam: PF01443; Viral_helicase1.
 DR Pfam: PF01660; Vmethyltransferase; 1.
 KW Transferrase: RNA-directed RNA polymerase: Helicase; ATP-binding.
 FT CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
 FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
 FT NP_BIND 833 840 ATP (POTENTIAL).
 SQ SEQUENCE 1616 AA; 183542 MW; 5DBFB2FADCC5C0C CRC64;

Query Match 8.0%; Score 98.5; DB 1; Length 1616;
 Best Local Similarity 20.4%; Pred. No. 16;
 Matches 40; Conservative 44; Mismatches 79; Indels 33; Gaps 7;

QY 57 KLVDLGLVQKSSAQLAKQAQAEKLNINVISKLLPTNTDIFGLKISNLSILDVKAEP 116
 Db 536 KMSVDMPLVD-----IRKXMEETEEMYNALSLSLVKSKDFDQVDFVFSOMCOSLEVD 588
 QY 117 IDGK-----GLNLSF--PVTANVTVAGPIIGIINLKASLDLLTAVTETDPQT 164
 Db 589 MTAAKVIVAVMSNESGSLTTEQPTANVALA-----LQDSEKASDCAALVVTSDV 639
 QY 165 HOPVAVLGEASDPTIS--LSLLDKHSQLINKFVNSVINTLKVTSVSLQKEICPLIRI 222
 Db 640 EEP-SIKSMARGELQAGLSDQVPESSYTRSEIESLEQPHMATASLHKOMCSI--V 696
 QY 223 FIHSLDVNVIOQVDN 238
 Db 697 YTGPLEKVQOMKNFIDS 712

RESULT 14
 DNMI_YEAST
 ID DNMI_YEAST STANDARD: PRT; 757 AA.
 AC P54861;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE Dnam1-related protein DNMI (EC 3.6.1.50).
 GN DNMI OR YLL001W OR I1381.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=93348179; PubMed=7622557;
 RA Gammie A.E., Kurihara L.J., Vaillee R.B., Rose M.D.;
 RT "DNMI, a dnam1-related gene, participates in endosomal trafficking
 RT in yeast."
 RL J. Cell Biol. 130:553-566(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY23;
 RX MEDLINE=96405918; PubMed=8810043;
 RA Miosga T., Zimmermann F.K.;
 RT "Sequence analysis of the CEN12 region of saccharomyces cerevisiae on
 RT a 43.7 kb fragment of chromosome XII including an open reading frame
 RT homologous to the human cystic fibrosis transmembrane conductance
 RT regulator protein CFTR";
 RL Yeast 12:693-708(1996).

RN [3]
 RP SEQUENCE FROM N.A.
 RA Vandenbol M., Portetelle D., Hilger F.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases
 CC FUNCTION: MICROTUBULE-ASSOCIATED FORCE-PRODUCING PROTEIN THAT
 CC PARTICIPATES IN ENOCYPTOSIS. DOES NOT APPEAR TO PARTICIPATE IN
 CC SECRETION OR VACUOLAR PROTEIN SORTING.
 CC -!- CATALYTIC ACTIVITY: GTP + H(2)O = GDP + phosphate.
 CC -!- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: L40588; AAA99998.1; -
 DR EMBL: X91488; CA62769.1; -
 DR EMBL: 273106; CAA97444.1; -
 DR SGD: S0003924; DNMI.
 DR InterPro: IPR001401; Dynamin.
 DR InterPro: IPR003375; Dynamin_central.
 DR InterPro: IPR003130; GED.
 DR Pfam: PF00350; dynamin; 2.
 DR Pfam: PF01031; dynamin_2; 1.
 DR Pfam: PF02212; GED; 1.
 DR PRINTS: PR00195; DYNAMIN.
 DR SMART: SM00053; DYNC; 1.
 DR SMART: SM00302; GED; 1.
 DR PROSITE: PS00410; DYNAMIN; 1.
 KW Hydrolase; Motor protein; GTP-binding.
 FT NP_BIND 35 42 GTP (POTENTIAL).
 FT NP_BIND 175 179 GTP (POTENTIAL).
 FT NP_BIND 244 247 GTP (POTENTIAL).
 FT CONFLICT 124 124 H -> ISPD (IN REF. 1).
 SQ SEQUENCE 757 AA; 84971 MW; EBEF8793C5951770 CRC64;

Query Match 7.9%; Score 97; DB 1; Length 757;
 Best Local Similarity 23.1%; Pred. No. 7.9;
 Matches 52; Conservative 35; Mismatches 74; Indels 64; Gaps 10;

QY 18 SESL-----LDNLGNDLSNVVDKLEPVLHGLETVLTKGILEKLVKD-LGVLOKSSAW 71
 Db 223 SESLKLAREVDPQGRRTIGVITKLD-LMDSGTNALD-ILSGKMYPLKLGFGVGVNRS--- 277
 QY 72 QLAQKAQAEKLLNNVISKLLPTNTDIP-----GLKISNLSILDVKAEPID 118
 Db 278 ---QDDIQ-----LAKTVEESLDKEDYERKHPVYRTISTCKGTRYLAKLL----- 320
 QY 119 DGKGINLSFPTANVTVAGPIIGIINLKASLDLLTAVTETDPQTHOPVAVLGEASD 178
 Db 321 -----NOTLLSHTRDKLPDIKTLNLTIS-----QTEQLARYGGVGATT 360
 QY 179 TSIISLLDKHSQLINKFVNSVINTLKVTSVSLQKEICPLIRIP 223
 Db 361 NESRASLV---LQLMKNFSTNPFISSIDGTSSDINTKELCGGARY 402

RESULT 15
 YGX7_YEAST
 ID YGX7_YEAST STANDARD: PRT; 958 AA.
 AC P53076;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE Hypothetical 108.2 kDa protein in SAP4-OST5 intergenic region.
 GN YGL227W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

```

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Fartmann B., Kramer B., Kramer W.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; 272749; CAA96943.1; .
DR SGD; S0003196; VID30.
DR InterPro; IPR003877; SPKY.
DR InterPro; IPR003878; SPKY_domain.
DR Pfam; PF00622; SPKY; 1.
DR SMART; SM00449; SPKY; 1.
KW Hypothetical protein.
FT DOMAIN 165 185 POLY-ASP.
FT DOMAIN 450 453 POLY-SER.
SQ SEQUENCE 958 AA; 108178 MW; 335ADD152949F8C8 CRC64;

Query Match 7.9%; Score 97; DB 1; Length 958;
Best Local Similarity 22.7%; Pred. No. 11;
Matches 58; Conservative 46; Mismatches 84; Indels 68; Gaps 13;

QY 12 GVLTGTSLSLDNLGND---LSNVVDKL--EPVLHGLETVDTNLTGILEKIKVDL---- 52
DB 690 GKLVDPDNNINLNSVDDGSLPNTLAWMINVDVLIHGLVDV---AKGFLKDLQKDAVNVN 746
QY 63 -----GVLOKSSANQLAKQK-----AQEAELLN-----NVISKLLPTNTD 98
DB 747 QHSESKDVIHNER-OIMKEERKWKIRGELRYLINKGOISKICINYIDNEIPDLLKNLE 805
QY 99 -IFGLKISNLIIDVKAEPIDG-----KGLNLS-----FPVTANVTYAGPIIQIINL 146
DB 806 LVPELKALYLVMIKKSSSKDDDEIENLILKGOELSNFIDYTKIPQSLDRFSGQLSNV 865
QY 147 KASLDLLTAVTITFDPTQHPVAVLGEASD-----PTSISLSLLDKHSQIINKFV 197
DB 866 SALL-----AYSNPLVEAPKEISGLSDEYLOERLFOVSNNITLTFLHKDSECA---L 915
QY 198 NSVINTLKSTVSSLIQ 213
DB 916 ENVISNTRAMLSLILE 931
```

Search completed: August 6, 2002, 17:07:31
Job time: 873 sec

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 16:53:58 : Search time 111.35 Seconds
(without alignments)
386.850 Million cell updates/sec

Title: US-10-020-139-2

Perfect score: 1233

Sequence: 1 MLQKWLKVLGCVLTGTSES.....NVIOQVVDNPOHKTQLTOLI 249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1233	100.0	249	4 Q9BQ00	Q9BQ00 homo sapien
2	1225	99.4	249	4 Q9BDR5	Q9BDR5 homo sapien
3	442	35.8	243	6 P79124	P79124 bos taurus
4	420.5	34.1	240	6 P79125	P79125 bos taurus
5	357.5	28.0	235	11 Q9D734	Q9D734 mus musculus
6	336.5	27.3	235	11 Q63471	Q63471 rattus norv
7	212.5	17.2	206	11 Q63550	Q63550 rattus norv
8	159	12.9	484	4 Q96HK6	Q96HK6 homo sapien
9	150.5	12.2	474	11 Q61114	Q61114 mus musculus
10	148.5	12.0	235	11 Q9D6P2	Q9D6P2 mus musculus
11	148.5	12.0	270	11 Q9D794	Q9D794 mus musculus
12	148.5	12.0	270	11 Q9C0X3	Q9C0X3 mus musculus
13	145.5	11.8	270	11 Q9D6P0	Q9D6P0 mus musculus
14	132	10.7	232	11 Q9D9J8	Q9D9J8 mus musculus
15	129	10.5	199	4 Q9BQ08	Q9BQ08 homo sapien
16	110.5	9.0	1075	10 Q9SMV6	Q9SMV6 arabidopsis

ALIGNMENTS

RESULT 1
Q9BQ00
ID Q9BQ00 PRELIMINARY: PRT: 249 AA.
AC Q9BQ00;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE BA49G10.1 (SIMILAR TO BOVINE SALIVARY PROTEIN BSP30).
GN BA49G10.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 111
RP SEQUENCE FROM N.A.
RA Tracey A.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL121901; CAC03546.1; -;
SQ SEQUENCE 249 AA; 27011 MW; E64E0794A1B4DB7D CRC64;

Query Match 100.0%; Score 1233; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 2e-82;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLQKWLKVLGCVLTGTSESLLDNLGNSLVYVDKLEPVLHGLETVDTNLTGILEKLV 60
|||||
Db 1 MLQKWLKVLGCVLTGTSESLLDNLGNSLVYVDKLEPVLHGLETVDTNLTGILEKLV 60
|||||
QY 61 DLGVLOKSSAWQLAKQKAGEAEKLLNNVSKLLPTNTDIFGLKISNLSLDVKAEPIDDG 120
|||||
Db 61 DLGVLOKSSAWQLAKQKAGEAEKLLNNVSKLLPTNTDIFGLKISNLSLDVKAEPIDDG 120
|||||
QY 121 KGLNLSFPVTANTVAGPIIGIQTINLKASLDLTAVTETDPTQHPVAVLGECAASDPTS 180
|||||
Db 121 KGLNLSFPVTANTVAGPIIGIQTINLKASLDLTAVTETDPTQHPVAVLGECAASDPTS 180
|||||
QY 181 ISLSLDKHSQIINKFVNSVINTLSTVSSLLQKELCPILRIFHSLDVNIQQVVDNPQ 240
|||||

Q99tn0 staphylococ
Q05701 rattus ratt
Q99r31 staphylococ
Q9a178 streptococ
Q9pf62 xyella fas
Q77136 apis mellif
Q54436 staphylothe
Q9cfl1 lactococcus
Q9c2c8 neurospora
Q99it0 tomato mosa
Q61352 mus musculu
Q9bq9 homo sapien
Q59040 methanococ
Q9kpm2 vibrio chol
Q9xtu2 caenorhabd
Q9wnq5 tobacco mos
Q9ja03 tobacco mos
Q9ja04 tobacco mos
Q9bjy0 plasmodium
Q61351 mus musculu
Q49348 mycoplasma
Q9dnt4 porcine rub
Q61769 mus musculu
Q48995 mycoplasma
Q9n218 caenorhabd
Q9ks47 vibrio chol
Q9cka9 pasteurilla
Q9fx8 oryza sativ
Q99u54 staphylococ

Db 181 ISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDYNVVIQOVVDNPO 240
Qy 241 HKTQLOTLI 249
Db 241 HKTQLOTLI 249

RESULT 2
Q96DR5 PRELIMINARY; PRT; 249 AA.
AC Q96DR5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (T-EMBlrel. 19, Last annotation update)
DE PAROTID SECRETORY PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PAROTID;
RA Venkatesh S.G., Gestha C., Gorr S.-U.;
RT "A member of the PSP/plunc family of BPI proteins is expressed in the
human parotid gland."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF432917; AAL28113.1;
SQ SEQUENCE 249 AA; 27110 MW; FD54B624A1A4CA7C CRC64;

Query Match 99.4%; Score 1225; DB 4; Length 249;
Best Local Similarity 99.6%; Pred. No. 7.8e-82;
Matches 248; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MLQWKLVLLCGVLTGTSESLLDNLGNDLSNVYDKLEPVLHGLETVDTNLTGILEKLKV 60
Db 1 MLQWKLVLLCGVLTGTSESLLDNLGNDLSNVYDKLEPVLHGLETVDTNLTGILEKLKV 60
Qy 61 DLGVLOKSSAWQLAKQAKAEKILNNVSKLLPTNTDIFGLKISNLSLIDVKAEPIDDG 120
Db 61 DLGVLOKSSAWQLAKQAKAEKILNNVSKLLPTNTDIFGLKISNLSLIDVKAEPIDDG 120
Qy 121 KGLNLSFPVTANVTAGPIIGQIINKASLDLLTAVTIEDTPQTHQPVAVLGECA SDPTS 180
Db 121 KGLNLSFPVTANVTAGPIIGQIINKASLDLLTAVTIEDTPQTHQPVAVLGECA SDPTS 180
Qy 181 ISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDYNVVIQOVVDNPO 240
Db 181 ISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDYNVVIQOVVDNPO 240
Qy 241 HKTQLOTLI 249
Db 241 HKTQLOTLI 249

RESULT 3
P79124 PRELIMINARY; PRT; 243 AA.
AC P79124;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE BSP30.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PAROTID GLAND;
RA Haigh B.J., Wilkins R.J., Wheeler T.T.;
RT "The cloning and sequencing of two cDNAs coding for alternate forms of
BSP30, a bovine member of the Parotid Secretory Protein family."
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U79414; AAB38283.1;
SQ SEQUENCE 240 AA; 26513 MW; 850611DE9E43E358 CRC64;

Query Match 34.1%; Score 420.5; DB 6; Length 240;
Best Local Similarity 38.1%; Pred. No. 3.5e-23;
Matches 93; Conservative 56; Mismatches 84; Indels 11; Gaps 5;

Qy 1 MLQWKLVLLCGVLTGTSESLLDNLGNDLSNVYDKLEPVLHGLETVDTNLTGILEKLKV 60
Db 1 MLQWKLVLLCGVLTGTSESLLDNLGNDLSNVYDKLEPVLHGLETVDTNLTGILEKLKV 60
Qy 61 DLGVLOKSSAWQLAKQAKAEKILNNVSKLLPTNTDIFGLKISNLSLIDVKAEPIDDG 120
Db 61 DLGVLOKSSAWQLAKQAKAEKILNNVSKLLPTNTDIFGLKISNLSLIDVKAEPIDDG 120
Qy 58 E---LESRCSDVEVEQ--QETENFEQLISRFQVSVKLTGVIRNVQVDPITFEATSEN 112
Db 58 E---LESRCSDVEVEQ--QETENFEQLISRFQVSVKLTGVIRNVQVDPITFEATSEN 112
Qy 121 KGLNLSFPVTANVTAGPIIGQIINKASLDLLTAVTIEDTPQTHQPVAVLGECA SDPTS 180
Db 121 KGLNLSFPVTANVTAGPIIGQIINKASLDLLTAVTIEDTPQTHQPVAVLGECA SDPTS 180
Qy 113 SA-NVLIPIADVTVSLPFLGEIVDLNVDLQTTVSIEDT--TEDPQVYVGECTNNPES 169
Db 113 SA-NVLIPIADVTVSLPFLGEIVDLNVDLQTTVSIEDT--TEDPQVYVGECTNNPES 169
Qy 181 ISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDYNVVIQOVVDNPO 240
Db 181 ISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDYNVVIQOVVDNPO 240

[illegible]

RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE-SUBMANDIBULAR GLAND;
RX MEDLINE=92129360; PubMed=1370829;
RA Mirels L., Ball W.D.;
RT "Neonatal rat submandibular gland protein SMG-A and parotid secretory
protein are alternatively regulated members of a salivary protein
multigene family.";
RL J. Biol. Chem. 267:2679-2687(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE-SUBMANDIBULAR GLAND;
RX MEDLINE=98129760;
RA Mirels L., Miranda A.J., Ball W.D.;
RT "Characterization of the rat salivary-gland B1-immunoreactive
proteins.";
RL Biochem. J. 330:437-444(1998).
DR EMBL; M83210; AAC12783.1;
KW Signal.
FT SIGNAL
FT CHAIN
FT FT
SQ SEQUENCE 206 AA; 23027 MW; 91C001620761067B CRC64;

Query Match 17.2%; Score 212.5; DB 11; Length 206;
Best Local Similarity 24.2%; Pred. No. 4.2e-08;
Matches 59; Conservative 53; Mismatches 85; Indels 47; Gaps 4;
QY 1 MLQWLKLVLCGVLGTSTSESLDNLGN---DLSNVVDKLEPVVHLEGTVD-NLKGILE 56
DB 1 MFQGLSVLVLCGLLGTSGSLFDIFQNPEDVESVNSEINRYRYALETMDLMDLADYLS 60
QY 57 KLVVDLGVLOKSSAWOLAKQKAOEAEKLLNNVSKLLPTNTDIFGLKISNSLIDVKAEP 116
DB 61 KRGLIE-----LKKDLRLNLNHEV 80
QY 117 IDGKGLNLSFPVTANTVAGPIIGQIINLKASLDLLTAVTETDPTQHPVAVLGEAS 176
DB 81 SPNGDEVTLKMPALNASLSLPARDLTVDTSISMEAITSPAIEKDPKTRVRLNMQRSL 140
QY 177 DPTSLSLDKHQSQINKFVNSVINTLKSTVSSLLQKEICPLRIFIHSLDNNVQVAV 236
DB 141 NFDNTSLSLNRKSNFVNIALDSALYLIKRGTLPLVRRQLCPVLQLIISN---TFHPDEI 197
QY 237 DNPQ 240
DB 198 SNPQ 201

RESULT 8
Q96HK6 PRELIMINARY; PRT; 484 AA.
AC Q96HK6
DT 01-DEC-2001 (TREMHLrel. 19, Created)
DT 01-DEC-2001 (TREMHLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMHLrel. 19, Last annotation update)
DE SIMILAR TO DNA SEGMENT, CHR 2, MASSACHUSETTS INSTITUTE OF
DE TECHNOLOGY 19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008429; AAH08429.1;
SQ SEQUENCE 484 AA; 52427 MW; 08242B697284E858 CRC64;
Query Match 12.9%; Score 159; DB 4; Length 484;
Best Local Similarity 23.9%; Pred. No. 0.00099;

Matches 58; Conservative 48; Mismatches 97; Indels 40; Gaps 9;
QY 5 WKLVLCGVLGTSTSESLDNLGNLSNVVDKLEP---VLHLEGTVDNTLKGILEKIVDL 62
DB 5 WPTLLCGLLAATL-----IOATLSPTAVLILG-----PKVKEKIQEIL 44
QY 63 -----GVLOKSSAWOLAKQKAOEAEKLLNNVSKLLPTNTDIFGLKISNSLIDVKAEP 116
DB 45 KHNATSILOQLPLLSAMREKPKAGGIPVGLSLVNPVL---KHVWLVKVTANILQVQVKP 101
QY 117 IDGKGLNLSFPVTANTVAGPIIGQIINLKASLDLLTAVTETDPTQHPVAVLGEAS 172
DB 102 SANQOELLVKVPLD---WVAGENTPIVKTIIVEFHMTE---AQATIMDTISASGPTKIVLS 156
QY 173 ECASDPTSLSLDKHQSQINKFVNSVINTLKSTVSSLLQKEICPLRIFIHSLDNNV 232
DB 157 DCATSHGSLKIQULHKLFLYNALAKOVNLLVPSLPNVLKQLCVPVIEASFGNYAVLL 216
QY 233 QVQ 235
DB 217 QLV 219

RESULT 9
Q61114 PRELIMINARY; PRT; 474 AA.
AC Q61114
DT 01-NOV-1996 (TREMHLrel. 01, Created)
DT 01-MAR-2001 (TREMHLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMHLrel. 17, Last annotation update)
DE VON EBER MINOR SALIVARY GLAND PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS-WEBSTER;
RA Sneed M.L., Villanueva J., Paine M.L., Lei Y.P., Zhu D.H., Luis J.,
Xia Y.-R., Yang J.-N.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U46068; AAA87581.2;
DR InterPro; IPR001124; LBP_BPI_CETP.
DR SMART; SM00328; BPI1; 1.
DR SMART; SM00329; BPI2; 1.
SQ SEQUENCE 474 AA; 52373 MW; 377DBFA5E736709B CRC64;

Query Match 12.2%; Score 150.5; DB 11; Length 474;
Best Local Similarity 23.1%; Pred. No. 0.004;
Matches 52; Conservative 56; Mismatches 84; Indels 33; Gaps 9;
QY 5 WKLVLCGVLGTSTSESLDNLGNLSNVVDKLEPVVHLEGTVDNTLKGILEK 57
DB 5 WPTLLCGLLGATLVQANVYPPAVL-NLGPEV-----IQKHLQALKHDAAT--ALQEE 55
QY 58 LKVDLGVLOKSSAWOLAKQKAOEAEKLLNNVSKLLPTNTDIFGLKISNSLIDVKAEP 117
DB 56 LPL-URAMQDKSG-----SIPILDFEVHTVLKY---ILWVKVTSANILQLDVQPS 101
QY 118 IDGKGLNLSFPVTANTVAGPIIGQIINLKASLDLLTAVTETDPTQHPVAVLGEAS 176
DB 102 TYNQELVWRIPLDVWAGLNTPLIKTIVEFQMSVQALIRVERSKS--GPAHLNLSQSS 159
QY 177 DPTSLSLDKHQSQINKFVNSVINTLKSTVSSLLQKEICPLIR 221
DB 160 NESTLRSLSLHKLKSLFVNSLAKVNNLLVPALPQIVKNNHLCPLMIQ 204
RESULT 10
Q9D6P2 PRELIMINARY; PRT; 235 AA.
ID Q9D6P2
AC Q9D6P2;

RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA	Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA	Sakai K., Okido T., Furuno M., Anon H., Baldarelli R., Barsh G.,
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA	Gustincich S., Hill D., Hofmann J., Hume D.A., Kamiya M., Lee N.H.,
RA	Lions P., Marchionni L., Mashima J., Mazzarelli J., Mombearts P.,
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA	Hayashtzaki Y.,
RA	*Functional annotation of a full-length mouse cDNA collection.*;
RA	Nature 409:685-690(2001).
RA	EMBL; AK009441; BAB26290.1; -.
DR	MGI; 1914385; 2310021H06Rik.
DR	SEQUENCE 270 AA; 29223 MW; DB63327561D910AA CRA64;
QY	Query Match 12.0%; Score 148.5; DB 11; Length 270;
Db	Best Local Similarity 25.3%; Pred. No. 0.0026;
QY	Matches 39; Conservative 38; Mismatches 74; Indels 3; Gaps
QY	84 LLNNVISKLLPTNTDFGLKINSLSILDVKAEPIDDKGMLNSLFPFYANTVAGPIIGQI 143
Db	: : : : :
QY	108 VLGKVSYS-IPLNLLINDIRVTNPQLLEIGLVQSVDHFRLVTVTFGLGDLRVNTLVVGS 166
QY	144 INLKASLDLITAVTIFETPOTHPVAVIGEACDSPTSISLSDKKHSQIIKNKFSVINT 203
Db	: : : : :
QY	167 LELSVKLDVTAENVVAVRDYSVGRSL-VIGDCITYPGSLURISLLNRGLPON-LIDSUTDI 224
QY	204 LKSTVSSLLKEICPLIRIFIHSLDVNVIOQVVD 237
Db	: : : : :
QY	225 LTRVPGLGVQVCPLVNGVLSLDTVLAHVDAD 258
QY	: : : : :
RESULT	12
QCXQ3	PRELIMINARY; PRT; 270 AA.
AD	QCXQ3 PRELIMINARY; PRT; 270 AA.
QCXQ3	QCXQ3:
DC	01-JUN-2001 (TrEMBRel. 17, Created)
DT	01-JUN-2001 (TrEMBRel. 17, Last sequence update)
DT	01-JUN-2001 (TrEMBRel. 17, Last annotation update)
DE	2310021H06RIK PROTEIN.
DE	2310021H06RIK.
GN	2310021H06RIK.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.
NCBI_Taxid	10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=TONGUE;
RX	MEDLINE=21085660; PubMed=11217851;
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA	Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA	Sakai K., Okido T., Furuno M., Anon H., Baldarelli R., Barsh G.,
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA	Lions P., Marchionni L., Mashima J., Mazzarelli J., Mombearts P.,
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

Query Match	11.8%	Score 145.5	DB 11	Length 270
Best Local Similarity	25.3%	Pred. No. 0.0046		
Matches 39	Conservative 38	Mismatches 74	Indels 3	Gaps
84	LNNVISKLLPTDITDFGLKINSLLIUVKAEIPDDGKGLNLSFPVTANVTYVAGPIIGOI	143		
108	VLGVVISI-IPLNLLNILDITRVNPNQLLEIGLVQSDYFRLVYVTIPGFDURVNTLVGSI	166		
144	INLKASLDLLTAVTIEDTPOTHQPAVLGECASOPTSISLSLKKHSIQINKEFVNSVT	203		
167	LELSVKLDVTAEVYAVRDFYGRSRL-VIGDCIYPPGSLKLSLLNRLGQLQN-LIDSUTDI	224		
204	LKSTVSSLLQKEICPLIRIPTHSLDNVNIQVWD	237		
225	LTRVIPGLVGQVCPVLNGVLSLLDYLTAHDVAD	258		
RESULT 14				
ID	Q9D9J8	PRELIMINARY;	PRT;	232 AA.
AC	Q9D9J8;			
DT	01-JUN-2001 (TReMBUrel. 17, Created)			
DT	01-JUN-2001 (TReMBUrel. 17, Last sequence update)			
DT	01-JUN-2001 (TReMBUrel. 17, Last annotation update)			
DE	1700058C13RIK PROTEIN.			
OS	1700058C13RIK.			
GN	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
ON	NCBI_TaxId=10090;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;			
RC	MEDLINE=21083660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavani T.,			
RA	Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,			
RA	Sakari L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Pronstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustineich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA	Lyonis P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA	Nordone P., King B., Ringwald M., Rodriguez L., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,			
RA	Wynshaw-Borisi A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,			
RA	Hayashizaki Y.			
RT	"Functional annotation of a full-length mouse cDNA collection."			
BL	Nature 409:685-690(2001).			
DR	ENBL; AK006829; BAB24760.1; -			
DR	MGI; 1920638; 1700058C13RIK.			
SQ	SEQUENCE 232 AA; 25713 MW; OD52D24A3076D5DC CRC64;			
Query Match	10.7%	Score 132	DB 11	Length 232
Best Local Similarity	21.18	Pred. No. 0.036		
Matches 55	Conservative	Mismatches 88	Indels 64	Gaps
QY	4 LWK-IYVLCGVLT-----GTSSLNDNLGNLSDNNVDKLEPVLHGLETVNTL	51		
DB	4 LWRVLLGLLGLALPSALPKPQWPLGTRKHKDGRST-----LARIAGCLKLN--A	53		

QY 52 KG-----ILEKLVLDLGVLOKSSAWQAKQAQAEKLLNNVSKLLPTNTDIFGLKIS 105
Db 54 EGRIQSMRLDLRLNSCTVAPCWGNLIGCMNFQQQOEISINI-----TNVL----- 101
QY 106 NSLILDKAEPIDDDGKGLNLSPP---VTANVTAG-----PIQOIINLKASLDLLTAV 156
Db 102 -----DCGGIQWAFPKWFSAITLFDIEFKLPFNSNIITKTHACMGLTAES 148
QY 157 TIETDPTQHPVAVIGECASDPTSISLSL-LDKHSQIINKFPVNSVINTLKSTVSSLSQKE 215
Db 149 WLEKDEFGREL--VMGRCRMEFPSSGCSMSTETSPKMKHFLHNLRESLKGKVPNLVESQ 207
QY 216 ICPLIRIFIHSLDVNVYVQV 236
Db 208 VCPICEILLQDLVDKLLGLV 228

RESULT 15
Q9BQP8 PRELIMINARY: PRT: 199 AA.
AC Q9BQP8;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE BA49G10.6 (SIMILAR TO MURINE VON EBNER MINOR SALIVARY GLAND PROTEIN,
DE ISOFORM 1) (FRAGMENT).
GN DJ1187J4.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tracey A.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL121901; CAC03550.1;
FT NON_TER 199
SQ SEQUENCE 199 AA; 21517 MW; 49A4CC2143BE04B1 CRC64;

Query Match 10.5%; Score 129; DB 4; Length 199;
Best Local Similarity 24.2%; Pred. No. 0.049;
Matches 54; Conservative 40; Mismatches 89; Indels 40; Gaps 9;
Qy 5 WKLVLCGLVTGTSESLDNLGNDLSNVKLEP--VLHECLETVONTLKGILEKLVKVDL 62
Db 5 WFTLLCGLLAATL-----IQATLSPTAVLIIG-----PKVIKEKLTQEL 44
Qy 63 -----GVLOKSSAWQAKQAQAEKLLNNVSKLLPTNTDIFGLKISNSLILDKAEP 116
Db 45 KDHNATSILOQLPLLSAMREKPGGIPVLGSLVNTVL---KHITWLKVTITANILQIQVKP 101
Qy 117 IDDGKGLNLSFPVTANVTAG---PIIQOIINLKASLDLLTAVTIEPTDPTQHPV-AVLG 172
Db 102 SANDOELLVKIPLD---MWAGFNTPLVKTIVEFHMTE--AQATIRMDTSASGPTRLVLS 156
Qy 173 ECASDPTSISLSDKHQSIINKFPVNSVINTLKSTVSSLSQKE 215
Db 157 DCATSHGSLRIQLLHLKLSFLNALAKQVMNLLVPSLPNLVKNQ 199

Search completed: August 6, 2002, 17:09:28
Job time: 930 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 16:47:18 : Search time 138.55 Seconds
(without alignments)
199.620 Million cell updates/sec

Title: US-10-020-139-2

Perfect score: 1233

Sequence: 1 MQLWKLVLLCGVLGTGTS...NVIOQVNDPNQKTLQTLI 249

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1233	100.0	249	19 AAW69221	Human parotid secr
2	1233	100.0	249	19 AAW60682	Human parotid secr
3	1233	100.0	249	21 AAB24069	Human PRO1025 prot
4	1233	100.0	249	21 AAB25765	Human secreted pro
5	1233	100.0	249	22 AAB75351	Human secreted pro
6	1195	96.9	260	22 AAB25745	Human protein sequ
7	246	20.0	50	22 ABB41435	Peptide #8941 enco
8	246	20.0	50	22 AAM62308	Human brain expres
9	246	20.0	50	22 AAM75111	Human bone marrow
10	246	20.0	50	22 AAM35227	Peptide #9264 enco
11	191.5	15.5	278	22 AAE05367	Mouse 28.6 kDa sec

12	183.5	14.9	256	20 AAY06408	Human secreted pro
13	183.5	14.9	256	20 AAW95463	LS170 polypeptide
14	183.5	14.9	256	21 AAY69164	Amino acid sequenc
15	183.5	14.9	256	21 AAY39019	Human secreted pro
16	183.5	14.9	256	22 AAY29210	Human PRO polypept
17	183.5	14.9	256	22 AAY39721	Human polypeptide
18	183.5	14.9	256	22 AAB97366	Human LUNX protein
19	183.5	14.9	256	22 AAY92209	Human digestive sy
20	183.5	14.9	264	22 AAM41507	Human polypeptide
21	168	13.6	187	22 AAM92212	Human digestive sy
22	167	13.5	191	22 AAM92214	Human digestive sy
23	166	13.5	320	22 AAM47214	Human digestive sy
24	160	13.0	318	22 AAM47220	Human NOV6a protei
25	160	13.0	484	21 AAY77126	Human NOV6b protei
26	180	13.0	565	22 ABG08520	Human neurotransmi
27	159	12.9	484	22 AAG63976	Novel human diagn
28	158	12.8	484	21 AAY99375	Amino acid sequenc
29	158	12.8	484	22 AAY29163	Human PRO1357 (UNQ
30	158	12.8	484	22 AAB87564	Human PRO polypept
31	158	12.8	484	22 AAB66124	Human PRO1357. Ho
32	122	9.9	197	21 AAY86219	Protein of the inv
33	122	9.9	221	21 AAB58378	Human secreted pro
34	108	8.8	751	22 AAU34336	Lung cancer associ
35	108	8.8	825	22 AAU37370	Staphylococcus aur
36	104	8.4	2710	17 AAR95016	Staphylococcus aur
37	104	8.4	2710	19 AAW68387	C. difficile toxin
38	102	8.3	2025	22 AAU34207	Clostridium diffic
39	102	8.3	3158	22 AAU37018	Staphylococcus aur
40	101	8.2	481	17 AAM16838	Staphylococcus aur
41	101	8.2	481	17 AAM16840	Recombinant endoto
42	100	8.1	1057	22 AAG67419	Amino acid sequenc
43	100	8.1	1057	22 AAB47212	Human KSP. Homo s
44	99	8.0	679	22 AAU35585	Haemophilus influe
45	98.5	8.0	210	20 AAY02695	Human secreted pro

ALIGNMENTS

RESULT 1

AAW69221

ID AAW69221 standard; Protein; 249 AA.

XX

AC AAW69221:

XX

DT 16-OCT-1998 (first entry)

XX

DE Human parotid secretory protein.

XX

KW Parotid secretory protein; hPSP, digestive disorder; endocrine disorder;

KW non-immune depressive disorder; immune system disorder; cancer; human;

KW therapy; diagnosis.

XX

OS Homo sapiens.

XX

FH Key

FT Peptide

FT Protein

FT

XX

PN W09828420-A1.

XX

PD 02-JUL-1998.

XX

PF 18-DEC-1997; 97WO-US23522.

XX

PR 23-DEC-1996; 96US-0034429.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Duan R, Ruben SM;

XX

Location/Qualifiers
1..18
/note= "signal peptide"
19..249
/note= "mature hPSP"

DR WPI: 1998-377651/32.
 DR N-PSDB; AAV44759.
 XX
 PT New nucleic acid encoding human parotid secretory protein or its
 PT fragments - useful for diagnosis and treatment of, e.g. digestive
 PT and endocrine disorders and for drug screening
 XX
 PS Claim 16; Fig 1: 94pp; English.
 XX
 CC This sequence is the human parotid secretory protein (hpsp) of the
 CC invention. The hpsp DNA is useful for chromosome identification and
 CC isolation of the corresponding genomic DNA. The DNA and protein can be
 CC used to detect abnormal levels of hpsp (in standard blotting,
 CC amplification or immuno assays), particularly for diagnosis of digestive,
 CC non-immune defensive, endocrine or immune system disorders. A particular
 CC application is diagnosis of cancers of the salivary gland, thymus and
 CC pancreas which are associated with high levels of hpsp. The protein is
 CC also useful as antifungal, antibacterial, antiparasitic and antiviral
 CC agents and may be expressed in vivo from the DNA. The protein, or cells
 CC expressing it, are used in screening tests to identify specific
 CC (antagonists, e.g. antibodies (Ab), polypeptides and antisense nucleic
 CC acids, which are potentially useful for treating conditions associated
 CC with excessive hpsp production. Cells containing the DNA are used to
 CC express the recombinant protein and this can be used to raise Ab, useful
 CC for diagnosis, therapy, for affinity purification and to identify
 CC hpsp-binding proteins.
 XX
 SQ Sequence 249 AA;
 0;
 Query Match 100.0%; Score 1233; DB 19; Length 249;
 Best Local Similarity 100.0%; Pred. No. 1.2e-100;
 Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLQLKLVLCGVLTGTSBLLDNLGNDLSNVVDKLEPVHLEGLETVNTLKGIKLV 60
 Db 1 mlqlkvlvllcgvltgtsesllndlgnldnsnvvdklepvihleglecvdntlkgielkv 60
 QY 61 DLGVLOKSSAWQLAKQAQAEKLLNNVSKLLPTNTDIFGLKISNSLILDVKAEPIDGG 120
 Db 61 dlgvloqssawqlakqaqaeekllnnvskllptntdfigikisnslildvkaepidgg 120
 QY 121 KGLNLSFPTANVTAGPIIGQIINLKASLDLLTAVTIETDPQHPVAVLGECDSDPTS 180
 Db 121 kglnsfpvtanvtvagpiigqilnnvskllptntdfigikisnslildvkaepidgg 180
 QY 181 ISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDNNVIOQVVDNQP 240
 Db 181 islsldkhsqiinkfvnsvintlkstvsllqkeicplirifihslldnnvliqqvvdnqp 240
 QY 241 HKTQLQTLI 249
 Db 241 hktqlqtl 249
 RESULT 2
 AAW60682
 ID AAW60682 standard; Protein; 249 AA.
 XX
 AC AAW60682;
 XX
 DT 18-SEP-1998 (first entry)
 XX
 DE Human parotid secretory protein (HSP).
 XX
 KW Parotid secretory protein; human; cancer; autoimmune disease;
 KW secretory tissue; gastrointestinal tissue; HSP; Sjorgen's syndrome;
 KW Graves disease; thyroiditis; insulin-dependent diabetes; pancreatitis;
 KW ulcerative colitis; Crohn's disease; atrophic gastritis.
 XX
 OS Homo sapiens.
 XX
 PN W09821329-A1.

XX 22-MAY-1998.
 XX
 PF 07-NOV-1997; 97WO-US20651.
 XX
 PR 14-NOV-1996; 96US-0749288.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Bandman O, Goli SK;
 XX
 DR WPI: 1998-297933/26.
 DR N-PSDB; AAV37699.
 XX
 PT New parotid secretory protein - useful for, e.g. treatment of cancer
 PT and auto-immune disease, particularly of secretory or
 PT gastrointestinal tissues
 XX
 PS Claim 1; Fig 1A-C; 65pp; English.
 CC This represents a human parotid secretory protein (HSP). Antagonists
 CC that bind specifically to, and modulate activity of HSP are used to
 CC treat cancer and autoimmune diseases particularly of secretory or
 CC gastrointestinal tissue, e.g. cancer of salivary gland, thyroid,
 CC prostate, breast, gastrointestinal tract or pancreas, Sjorgen's syndrome,
 CC Graves disease, thyroiditis, insulin-dependent diabetes, pancreatitis,
 CC ulcerative colitis, Crohn's disease and atrophic gastritis. Host cells
 CC containing expression vectors comprising the HSP nucleic acid are used
 CC to produce recombinant HSP which is used to generate antibodies and to
 CC screen for its antagonists. Antibodies are useful directly as
 CC antagonists, to transport drugs to HSP-expressing cells, to detect cells
 CC that express HSP, to monitor patients being treated with HSP, and for
 CC purification of HSP from natural sources. Expression of HSP may
 CC indicate cell proliferation. HSP nucleic acid or its fragments are used
 CC to detect HSP-encoding sequences (optionally after amplification by PCR)
 CC by hybridisation, particularly for diagnosis and monitoring of disease,
 CC but also for mapping the chromosomal sequence.
 XX
 SQ Sequence 249 AA;
 0;
 Query Match 100.0%; Score 1233; DB 19; Length 249;
 Best Local Similarity 100.0%; Pred. No. 1.2e-100;
 Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLQLKLVLCGVLTGTSBLLDNLGNDLSNVVDKLEPVHLEGLETVNTLKGIKLV 60
 Db 1 mlqlkvlvllcgvltgtsesllndlgnldnsnvvdklepvihleglecvdntlkgielkv 60
 QY 61 DLGVLOKSSAWQLAKQAQAEKLLNNVSKLLPTNTDIFGLKISNSLILDVKAEPIDGG 120
 Db 61 dlgvloqssawqlakqaqaeekllnnvskllptntdfigikisnslildvkaepidgg 120
 QY 121 KGLNLSFPTANVTAGPIIGQIINLKASLDLLTAVTIETDPQHPVAVLGECDSDPTS 180
 Db 121 kglnsfpvtanvtvagpiigqilnnvskllptntdfigikisnslildvkaepidgg 180
 QY 181 ISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDNNVIOQVVDNQP 240
 Db 181 islsldkhsqiinkfvnsvintlkstvsllqkeicplirifihslldnnvliqqvvdnqp 240
 QY 241 HKTQLQTLI 249
 Db 241 hktqlqtl 249
 RESULT 3
 AAB24069
 ID AAB24069 standard; Protein; 249 AA.
 XX
 AC AAB24069;
 XX
 DT 29-JAN-2001 (first entry)

XX DE Human PRO1025 protein sequence SEQ ID NO:38.
 XX KW Human: tumour; diagnosis; neoplastic disease; neoplastic cell growth;
 KW proliferation; tumorigenesis; identification; cancer; cytostatic;
 KW neuroprotection; neuroprotective; anti-inflammatory; immunosuppressive;
 KW immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
 KW neuronal disorder; glioma disorder; astrocytic disorder; angiogenic;
 KW hypothalamic disorder; glandular disorder; macrophage disorder;
 KW epithelial disorder; stromal disorder; blastocoele disorder;
 KW inflammatory disorder; immunologic disorder.
 XX OS Homo sapiens.
 XX PN WC200053755-A2.
 XX PD 14-SEP-2000.
 XX PF 06-JAN-2000; 2000WO-US00376.
 XX PR 08-MAR-1999; 99WO-US05028.
 PR 02-JUN-1999; 99WO-US12252.
 PR 23-JUN-1999; 99US-0141037.
 PR 07-JUL-1999; 99US-0143048.
 PR 26-JUL-1999; 99US-0145698.
 PR 30-NOV-1999; 99WO-US28313.
 PR 20-DEC-1999; 99WO-US30911.
 PR 05-JAN-2000; 2000WO-US00219.
 XX PA (GETH) GENENTECH INC.
 XX PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;
 PI Watanabe CK, Wood WI;
 XX DR WPI; 2000-572270/53.
 DR N-PSDB; AAC58379.
 XX PT Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
 PT treatment, diagnosis and prevention of cancer -
 XX Claim 61; Fig 26; 286pp; English.
 XX The present invention describes an isolated antibody that binds to
 CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
 CC PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
 CC PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1182, PRO1184,
 CC PRO1187, PRO1281, PRO1283, PRO139, PRO134, PRO1317, PRO1710, PRO2094,
 CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
 CC growth. The PRO polypeptides and nucleotides are useful in the
 CC treatment, diagnosis and prevention of cancer. The antibodies and other
 CC anti-tumour compounds may be used to treat various conditions, including
 CC those characterised by overexpression and/or activation of the amplified
 CC PRO genes. Exemplary conditions or disorders to be treated with such
 CC antibodies and other compounds include benign or malignant tumours
 CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
 CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
 CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
 CC leukaemias and lymphoid malignancies, other disorders such as neuronal,
 CC glioma, astrocytic, hypothalamic and other glandular, macrophage,
 CC epithelial, stromal and blastocoele disorders, and inflammatory,
 CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
 CC primers and hybridisation probes used in the isolation of the human PRO
 CC sequences. AAC58367 to AAC58396 and AAC24057 to AAC24089 represent human
 CC PRO polynucleotide and protein sequences given in the exemplification of
 CC the present invention.
 XX Sequence 249 AA;

Query Match 100.0%; Score 1233; DB 21; Length 249;
 Best Local Similarity 100.0%; Pred. No. 1.2e-100;
 Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLCGLKVLICGVLTCTSESLLDNLGNDLSNVVDKLEPVHLEGLETVDNTLKGILEKLV 60
 Db 1 mlq;wklvllcgvltctsesllndlnldnsnvvdvklpvlhegletdvntlkgilekvl 60
 QY 61 DLGVLOKSSAWQAKQKQAEKLLNNVSKLLPTNTDIFGLKISNSLIDVKAEPIDG 120
 Db 61 dlgvloksawqakqkaeakllnnvskllptntdfigiknsllldvkaepidg 120
 QY 121 KGLNLSFPVTANTVAGPIIGQIINLKASLDLLTAVTIETDPQTHQPVAVLGECA SDPTS 180
 Db 121 kglnlspvtantvavagpiigqilnlkasldlltavti etdpqthqpvavlgecas dpts 180
 QY 181 ISLSLLDKHSQIINKEFVNSVINTKSTVSSLLQKEICPLIRIFIHSLDYNVIQQVVDNPQ 240
 Db 181 islslldkhsqiinkfenvsvintkstvssllqkeicplirifihsl dvnviqqvvdnpq 240
 QY 241 HKTQLQTLI 249
 Db 241 hktqlqtl i 249
 RESULT 4
 AAB25765
 ID AAB25765 standard; Protein; 249 AA.
 XX AC AAB25765;
 XX DT 28-NOV-2000 (first entry)
 XX DE Human secreted protein SEQ ID #77.
 XX KW Human: secreted protein; forensic procedure; gene therapy;
 KW chromosome mapping; cancer; autoimmune disease; cardiovascular disorder;
 KW cystic fibrosis; hypothyroidism; immunological disorder; amyloidosis;
 KW brain disorder; skeletal muscle disorder; eye disorder; obesity;
 KW mitochondrial cytopathy; diabetes; atherosclerosis; Alzheimer's disease;
 KW neurodegenerative disorder; graft rejection; dementia; hyperlipidaemia;
 KW septic shock; impotence.
 XX OS Homo sapiens.
 XX PN WO200037491-A2.
 XX PD 29-JUN-2000-
 XX PF 20-DEC-1999; 99WO-IB02058.
 PR 22-DEC-1998; 98US-0113686.
 PR 25-JUN-1999; 99US-0141032.
 XX (GETH) GENSET.
 XX Bougueleret L, Dumas J, Duclert A;
 WPI; 2000-442637/38.
 N-PSDB; AAA87727.
 XX Polynucleotides and polypeptides encoding proteins with signal
 PT peptides, useful in diagnostic, forensic, gene therapy and chromosome
 PT mapping procedures -
 XX Claim 9; Figure 10; 306pp; English.
 CC This sequence represents a human secreted protein amino acid sequence.
 CC The invention relates to sequences AAA87725-AAA87774 which encode human
 CC secreted proteins AAB25763-B25812. The proteins include signal peptides.
 CC Included in the invention are a host cell containing one of the cDNA
 CC sequences, and a purified antibody capable of binding to one of the
 CC secreted proteins. Also contained in the invention are methods for
 CC storing the sequence data on a computer system, and a method for
 CC identifying features of the cDNA sequences using a computer programme.
 CC The cDNAs are useful for expressing secreted proteins or fragments to
 CC obtain antibodies capable of specifically binding to the secreted

CC proteins. The CDNAS may also be useful in diagnostic, forensic, gene
 CC therapy and chromosome mapping procedures and may be used to design
 CC expression vectors and secretion vectors. The proteins of the invention
 CC may be used to treat diseases including cancer, autoimmune diseases,
 CC cardiovascular disorders, cystic fibrosis, hypothyroidism, immunological
 CC disorders, amyloidosis, brain disorders, skeletal muscle disorders, eye
 CC disorders, obesity, mitochondrial cytopathies, diabetes, atherosclerosis,
 CC neurodegenerative disorders, graft rejection, Alzheimer's disease,
 CC dementia, hyperlipidaemia, septic shock and impotence.

XX Sequence 249 AA;

Query Match 100.0%; Score 1233; DB 21; Length 249;
 Best Local Similarity 100.0%; Pred. No. 1.2e-100;
 Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLQIWKLVLCGVLTGTSESLLDNLGNLSNVVDKLEPVLHGLETVDTNLTGILEKLV 60
 Db 1 MLQIWKLVLCGVLTGTSESLLDNLGNLSNVVDKLEPVLHGLETVDTNLTGILEKLV 60
 Qy 61 DLGVLOKSSAWOLAKOQAQAEKLLNNVSKLLPTNTDIFGLKISNSLIDVKAEPIDDG 120
 Db 61 DLGVLOKSSAWOLAKOQAQAEKLLNNVSKLLPTNTDIFGLKISNSLIDVKAEPIDDG 120
 Qy 121 KGLNLSFPVTANVTVAGPIIGQIINLKASLDLLTAVTIEDTPQTHOPVAVLGECDSPTS 180
 Db 121 KGLNLSFPVTANVTVAGPIIGQIINLKASLDLLTAVTIEDTPQTHOPVAVLGECDSPTS 180
 Qy 181 ISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFHSLDYNVIOQVVDNPQ 240
 Db 181 ISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFHSLDYNVIOQVVDNPQ 240
 Qy 241 HKTQLQTLI 249
 Db 241 HKTQLQTLI 249

RESULT 5

ID AAB75351 standard; protein: 249 AA.

XX AAB75351;

DT 05-APR-2001 (first entry)

DE Human secreted protein #10.

XX Secreted protein; prevention; treatment; diagnosis; disease;
 XX infection.

XX Homo sapiens.

XX WO200100806-A2.

PD 04-JAN-2001.

PF 21-JUN-2000; 2000WO-1B00951.

PR 25-JUN-1999; 99US-0141032.

XX 21-DEC-1999; 99US-0469099.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Bougueleret L, Jobert S;

DR WPI; 2001-071487/08.

XX 49 Secreted proteins and the nucleic acids encoding them, useful in
 PT gene therapy and for detecting similar sequences in samples

PS Claim 10; Page 281; 307pp; English.

XX

CC The present invention relates to 49 Secreted proteins and the CDNAS
 CC encoding them. The protein and nucleic acids may be used in the
 CC prevention, treatment and diagnosis of diseases associated with
 CC inappropriate protein expression.

XX Sequence 249 AA;

Query Match 100.0%; Score 1233; DB 22; Length 249;
 Best Local Similarity 100.0%; Pred. No. 1.2e-100;
 Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLQIWKLVLCGVLTGTSESLLDNLGNLSNVVDKLEPVLHGLETVDTNLTGILEKLV 60
 Db 1 MLQIWKLVLCGVLTGTSESLLDNLGNLSNVVDKLEPVLHGLETVDTNLTGILEKLV 60
 Qy 61 DLGVLOKSSAWOLAKOQAQAEKLLNNVSKLLPTNTDIFGLKISNSLIDVKAEPIDDG 120
 Db 61 DLGVLOKSSAWOLAKOQAQAEKLLNNVSKLLPTNTDIFGLKISNSLIDVKAEPIDDG 120
 Qy 121 KGLNLSFPVTANVTVAGPIIGQIINLKASLDLLTAVTIEDTPQTHOPVAVLGECDSPTS 180
 Db 121 KGLNLSFPVTANVTVAGPIIGQIINLKASLDLLTAVTIEDTPQTHOPVAVLGECDSPTS 180
 Qy 181 ISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFHSLDYNVIOQVVDNPQ 240
 Db 181 ISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFHSLDYNVIOQVVDNPQ 240
 Qy 241 HKTQLQTLI 249
 Db 241 HKTQLQTLI 249

RESULT 6

ID AAM25745 standard; protein: 260 AA.

XX AAM25745;

DT 16-OCT-2001 (first entry)

DE Human protein sequence SEQ ID NO:1260.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiagregant; haemostatic; vulnery; antidiabetic; osteopathic; eczema;
 KW dermatological; antiallergic; antisthmatic; antiparkinsonian; infection;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.

XX Homo sapiens.

XX WO200153455-A2.

XX 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

PI Taij Yt, Liu C, Drmanac RT;
 DR WPI; 2001-457603/49.
 DR N-PSDB; AAH95686.
 XX
 PT Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
 XX
 PS Claim 20; Page 260; 1217pp; English.
 XX
 CC AAH9166, to AAH9904 encode the human proteins given in AAH25225 to
 CC AAH25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antinaemic; antiaggregant; haemostatic; vulnerary;
 CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytosstatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, haematopoietic disorders,
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.
 XX
 SQ Sequence 260 AA;
 Query Match 96.9%; Score 1195; DB 22; Length 260;
 Best Local Similarity 97.2%; Pred. No. 2.7e-97;
 Matches 242; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 QY 1 MLQWKIVLCGVLTGTSLLDNLGNLSNVVDKLEPVLHSGLETVDNTLKGILEKLV 60
 Db 12 mlglwklvllcgvltgtsesllldnlgndlsnvvdcklepvlhsgletvdtntlkgileklv 71
 QY 61 DLGVLOKSSAWOLAKQAEKLLNNVSKLLPTNTDIFGLKINSLLDVKAPIDDG 120
 Db 72 dlgyvkssawqlakqaeakllnnvskllptntdfigklnslldvkaepiddg 131
 QY 121 KGLNLSFPVTANYTVAGPIIGQIINLKASLDLLTAVTIETDPQTHOPVAVLGECDSDPTS 180
 Db 132 kglnlsfpvtanvtesqplldglinlralsldltavtietsdpqthhpvavlgcecardpts 191
 QY 181 ISLSLLDKHNSQIINKFVNSVINTLSTVSSILQKEICPLIRIFHSLDNNVTOQVVDNPQ 240
 Db 192 islsldkhsqinkfvnsvinclstvsllqkeicplirifihslndvnlqgvvdnpq 251
 QY 241 HKTQLOTLI 249
 Db 252 hktqlgtli 260
 RESULT 7
 ABB41435
 ID ABB41435 standard; Peptide: 50 AA.
 XX
 AC ABB41435;
 XX
 XX
 DT 04-FEB-2002 (first entry)
 XX
 DE Peptide #8941 encoded by human foetal liver single exon probe.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.
 XX
 OS Homo sapiens.

XX WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-483447/52.
 XX
 DR Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver.
 XX
 PS Claim 27; SEQ ID NO 34070; 639pp + sequence listing; English.
 XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 50 AA;
 Query Match 20.0%; Score 246; DB 22; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.2e-14;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 138 PIIGQIINKASLDLLTAVTIETDPQTHOPVAVLGECDSDPTSISLSLD 187
 Db 1 pligqinlkasldlltavietsdpqthpavlgceasdpstsisld 50
 RESULT 8
 AAM62308
 ID AAM62308 standard; Protein: 50 AA.
 XX
 AC AAM62308;
 XX
 DT 05-NOV-2001 (first entry)
 XX
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 34413.
 XX
 KW Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157275-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00667.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -

XX Example 4; SEQ ID NO: 34413; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.

XX Sequence 50 AA;

Query Match 20.0%; Score 246; DB 22; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.2e-14;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 138 PIIGCIINLKASLDLLTAVTETDPQTHQPVAVLGCASDPTISLSLLD 187
 Db 1 p1lgqiinkasidlltavltdpqtghpavlgcasdptsislid 50

RESULT 9

AAW75111
 ID AAW75111 standard; Protein; 50 AA.

XX AC AAW75111;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 35417.

XX Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312;

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -

XX

XX Example 4; SEQ ID NO: 35417; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.

XX Sequence 50 AA;

Query Match 20.0%; Score 246; DB 22; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.2e-14;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 138 PIIGCIINLKASLDLLTAVTETDPQTHQPVAVLGCASDPTISLSLLD 187
 Db 1 p1lgqiinkasidlltavltdpqtghpavlgcasdptsislid 50

RESULT 10

AAW35227
 ID AAW35227 standard; Protein; 50 AA.

XX AC AAW35227;

XX 17-OCT-2001 (first entry)

XX Peptide #9264 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -

XX Claim 27; SEQ ID NO 35496; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs;
 CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.

XX Sequence 50 AA;

Query Match 20.0%; Score 246; DB 22; Length 50;

CC encoding novel secreted proteins (see AAY06404-10) of the human
 CC testis, brain and foetal kidney. The polynucleotides and
 CC proteins are predicted to have biological activities which would
 CC make them suitable for treating, preventing or ameliorating medical
 CC conditions in humans and animals, although no supporting data are
 CC provided. Suggested activities include nutritional, cytokine, cell
 CC proliferation or differentiation, immunostimulant (e.g. as
 CC vaccine) or immunosuppressive, haematopoietic regulating, tissue
 CC growth, activin or inhibin; chemotactic or chemokinetic,
 CC haemostatic, thrombolytic, receptor/ligand, antiinflammatory,
 CC cadherin or tumour invasion suppressor, and tumour inhibition
 CC activity.

XX Sequence 256 AA;

Query Match 14.9%; Score 183.5; DB 20; Length 256;
 Best Local Similarity 23.8%; Pred. No. 3.4e-08;
 Matches 56; Conservative 46; Mismatches 100; Indels 33; Gaps 6;

QY 15 TGTSESLDNLGNDLSNVVDKLEPVVHLEGLETVDNTLKGILEKLVGLVQKSSAWOLA 74
 DB 45 TGLAGSLTNAISNGI-----ISGGL-----LGIENPLIDIL 77
 QY 75 KKAQAEKLLNNVISK---LLPTNTDIFGLKISNLSILDVKAEPIDGKGLNLSFPVTA 131
 DB 78 KPGGYSGLLGGILGKVSVIPGIMNIDIKVTDQQLLEIGVSPDGRHRYVTPLGI 137
 QY 132 NVTVAGPIIG-QIINKASDLTLTAVTITETDPQTHQPVAVLGECAEDPTISISLLDKHS 190
 DB 138 KQVNTPLVGSAILRAVKIDITAEILAVRDKQERHIL-VLGDCHSPGSIQISLIDGLG 196
 QY 191 QI-INKEFVNSVINTLKVSSLLQKEICPLIRIFIHSLDVNVVIOQVVDNPOHKTQ 244
 DB 197 PIPQGLIDSLTGILNKVLPVGVNCPVNEVIRGLDITLVHDMNMLIHGLQ 251

RESULT 13

AAW95463
 ID AAW95463 standard; Protein: 256 AA.

XX AC AAW95463;

XX DT 29-MAR-1999 (first entry)

XX DE LS170 polypeptide sequence.

XX KW LS170 gene; in vivo imaging; lung disease; cancer; drug screening.

XX OS Homo sapiens.

XX PN WO9856951-A1.

XX PD 17-DEC-1998.

XX PF 11-JUN-1998; 98WO-US11601.

XX PR 11-JUN-1997; 97US-0049183.

XX PA (ABBO) ABBOTT LAB.

XX PI Billing-medel PA, Cohen M, Colpitts TL, Friedman PN;
 Gordon J, Granados, Hodges SC, Klass MR, Kratochvil JD;
 Roberts-rapp L, Russell JC, Stroupe SD;

XX DR WPI; 1999-060335/05.

XX DR N-PSDB; AAX00809.

XX PT New LS170 nucleic acid from lung tissue - useful for detecting,
 PT monitoring, preventing and treating lung disease, particularly
 XX cancer

PS Claim 17; Page 96-97; 120pp; English.

XX

CC Sequences AAW95463-471 represent LS170 polypeptide fragments which
 CC contain at least one epitope in their sequences. The present invention
 CC relates to detection of a target LS170 polynucleotide that comprises
 CC treating a test sample with at least one LS170-specific nucleic acid that
 CC has at least 50 percent identity with any of the sequences (AAX00801-09)
 CC which are fragments derived from various clones of LS170 gene. The LS170
 CC nucleic acid fragments represent a set of contiguous, partially
 CC overlapping sequences transcribed from lung tissue. They are used for
 CC diagnosis, staging, monitoring, in vivo imaging, prevention and treatment
 CC of lung disease, specifically cancer, and to indicate predisposition to
 CC such disease. Particularly detection of LS170 polynucleotide, LS170
 CC antigens, or anti-LS170 antibodies is indicative of disease. Cells
 CC transformed with an expression system comprising the LS170 nucleic acid
 CC sequences are used to express recombinant polypeptides. The polypeptides
 CC are used to raise Ab and for drug screening. LS170-related nucleic acid
 CC can be used to isolate related sequences; as standards and reagents in
 CC assays; as targets for drug screening, and as components or targets for
 CC therapy, e.g. as antisense, ribozyme or triplex-forming agents. Ab can be
 CC used to deliver therapeutic agents to LS170-expressing cells. Directly
 CC as therapeutic agents (by neutralising LS170 polypeptides); in
 CC competitive binding drug screens, and to generate anti-idiotypic
 CC antibodies for use in rational drug design.

XX Sequence 256 AA;

Query Match

Best Local Similarity 14.9%; Score 183.5; DB 20; Length 256;
 Matches 56; Conservative 46; Mismatches 100; Indels 33; Gaps 6;

QY 15 TGTSESLDNLGNDLSNVVDKLEPVVHLEGLETVDNTLKGILEKLVGLVQKSSAWOLA 74
 DB 45 TGLAGSLTNAISNGI-----ISGGL-----LGIENPLIDIL 77
 QY 75 KKAQAEKLLNNVISK---LLPTNTDIFGLKISNLSILDVKAEPIDGKGLNLSFPVTA 131
 DB 78 KPGGYSGLLGGILGKVSVIPGIMNIDIKVTDQQLLEIGVSPDGRHRYVTPLGI 137
 QY 132 NVTVAGPIIG-QIINKASDLTLTAVTITETDPQTHQPVAVLGECAEDPTISISLLDKHS 190
 DB 138 KQVNTPLVGSAILRAVKIDITAEILAVRDKQERHIL-VLGDCHSPGSIQISLIDGLG 196
 QY 191 QI-INKEFVNSVINTLKVSSLLQKEICPLIRIFIHSLDVNVVIOQVVDNPOHKTQ 244
 DB 197 PIPQGLIDSLTGILNKVLPVGVNCPVNEVIRGLDITLVHDMNMLIHGLQ 251

RESULT 14

AAW95463
 ID AAW95463 standard; Protein: 256 AA.

XX AC AAW95463;

XX DT 30-MAY-2000 (first entry)

XX DE Amino acid sequence of lung specific protein Lngl10.

XX KW Lung specific gene; lung cancer; metastatic lung cancer; imaging.

XX OS Homo sapiens.

XX PN WO200008206-A1.

XX PD 17-FEB-2000.

XX PF 19-JUL-1999; 99WO-US16247.

XX PR 04-AUG-1998; 98US-0095233.

XX PA (DIAD-) DIADEXUS LLC.

XX PI Yang F, Sun Y, Recipon H, Macina RA;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 17:05:10 ; Search time 66.51 Seconds
(without alignments)
358.294 Million cell updates/sec

Title: US-10-020-139-2_COPY_2_249
Perfect score: 1228
Sequence: 1 LQLWKLVLGCVLTGTSES.....NVIQVVDNPNQKTLQTLI 248

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR71:*
1: Pirl:*
2: Pirl2:*
3: Pirl3:*
4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	359.5	29.3	235	1	SOMS	parotid secretory
2	331.5	27.0	235	2	B42337	parotid secretory
3	207.5	16.9	206	2	A42337	submandibular gland
4	110.5	9.0	1075	2	T52638	Exportin 1 (xpol) p
5	110.5	9.0	1075	2	T51557	Exportin 1 (xpol) p
6	109	8.9	1769	2	S53378	probable membrane
7	108	8.8	825	2	B89944	hypothetical prote
8	106	8.6	296	2	A49943	tructose-bisphosph
9	104.5	8.5	473	2	S17418	probable ligand-bl
10	104.5	8.5	1727	2	T30073	myosin-like coiled
11	104	8.5	1616	2	G42422	cytadherence-acces
12	104	8.5	2710	2	A37052	toxin A - Clostrid
13	103.5	8.4	296	2	F90067	hypothetical prote
14	100.5	8.2	1056	1	G02157	kinesin-like spind
15	100	8.1	990	2	C82759	zinc proteinase XF
16	99.5	8.1	1524	2	S68553	surface layer prot
17	99	8.1	679	2	D64182	DNA ligase (NAD+)
18	99	8.1	1441	2	B68807	hypothetical prote
19	98.5	8.0	1615	1	NMTMT8	180K protein - tom
20	98	8.0	521	2	S34338	biliary glycoprote
21	97.5	7.9	211	2	D64505	hypothetical prote
22	97.5	7.9	414	2	T27045	hypothetical prote
23	97	7.9	757	1	S64742	dynamitin-related pr
24	97	7.9	958	2	S64249	hypothetical prote
25	96.5	7.9	328	2	A82087	phosphoserine phos
26	96.5	7.9	1345	2	H90975	hypothetical prote
27	96.5	7.9	2660	2	E85822	probable invasiv z
28	96	7.8	458	2	JC1509	biliary glycoprote
29	96	7.8	624	2	PC6003	surface membrane p

30	96	7.8	1005	2	A54465	hypothetical prote
31	95.5	7.8	2938	2	T30249	cell proliferation
32	95	7.7	256	2	S77810	probable DNA topoi
33	95	7.7	578	2	B82204	methyl-accepting c
34	95	7.7	1130	2	T19148	hypothetical prote
35	95	7.7	6713	2	B89921	hypothetical prote
36	94.5	7.7	292	2	H89824	conserved hypothet
37	94.5	7.7	1729	2	S57596	ribosomal RNA proc
38	94	7.7	462	2	T10094	nitrogenase (EC 1.
39	94	7.7	490	2	D84999	ketol-acid reducto
40	93.5	7.6	640	2	T03754	hypothetical prote
41	93.5	7.6	722	2	H97217	uncharacterized co
42	93.5	7.6	868	2	S65186	NIP80 protein - ye
43	93.5	7.6	1038	2	S37854	hypothetical prote
44	93	7.6	206	2	A40305	biliary glycoprote
45	93	7.6	752	2	T09649	delta-1-pyrroline-

ALIGNMENTS

RESULT 1

SQMS

parotid secretory protein precursor - mouse

N: Alternate names: PSP

C: Species: Mus musculus (house mouse)

C: Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 22-Jun-1999

C: Accession: A23031: I53236

R: Madsen, H.O.; Horth, J.P.

13, 1-13, 1985

cloning of mouse PSP mRNA.

A23031: MUID:85215456

A

GB:X01697; NID:g53810; PIDN:CAA25846.1; PID:g759163

sen, B.K.; Mikkelsen, B.W.; Harmark, K.; Nielsen, J.T.; Hjorth, 1986

of murine parotid secretory protein and salivary amylase expre

I53236; MUID:87004556

; translated from GB/EMBL/DBDJ

cross-references: GB:M26807; NID:g200556; PIDN:AAA40009.1; PID:g554264

C: Comment: PSP is the most abundant protein in the parotid gland. Its function is n

C: Genet: A: Gene: Psp

A: Map position: 2

A: Introns: 41/1

A: Note: list of introns may be incomplete

C: Superfamily: parotid secretory protein

C: Keywords: parotid gland; saliva

F: 1-20/Domain: signal sequence #status predicted <STIG>

F: 21-235/Product: parotid secretory protein #status predicted <MAT>

Query Match

Best Local Similarity 29.3%; Score 359.5; DB 1; Length 235;

Matches 80; Conservative 62; Mismatches 79; Indels 17; Gaps 3;

QY 2 QLKWLVLGCVLTGTSESILNDNLGNDLNPVWVKLEPVLHGLETVNDTLKGILEKLVKL 61

DB 3 QLGSLVLGCVLTGTSESILNDNLGNDLNPVWVKLEPVLHGLETVNDTLKGILEKLVKL 50

QY 62 GVLQSSAQLAKQAQAEKLNINVKLLPTNTDIFG--LKISNLSILDVKAEPIDGG 119

DB 51 ELUQATSSPLAKNSILET---LNTADLGNLKSFTSLNGLLKNLNKLVDFQAKLSNG 107

QY 120 KGLNLSFPVTANVTAVAGPIIGIQLINLKASLDLTAVTIETDPTQHPVAVLGECASDPTS 179

DB 108 NGIDLTVPLAGEASLVLPFIKGTVDIVSILNLSIKTKNAQTGLPEVITGKSSWTDK 167

db	3	QUGSLVVLCCLLGTSGSLPDI	PQNPELDVESYVSEINRYRYALETMTDMLADYLSKR	62
QY	58	KVDGLVQLXSSAWOLAKQAQAEKLLNNV	ISKLLPTNTDIFGLKTSNLSILDVKAEPID	117
db	63	G1E-	-----LKIKDLRILNLNHEVSP	82
QY	118	DGRGLNLSFFVYANVTAVGPI	IGQIINLKASLDLITAVTITETDPTQHPVAVLGECA	177
db	83	NCIDEVTLKPMHMLNASLSLPARULDT	TVDSISMEAITSFALEKDPKTRRVNLNMQRCSLNT	142
QY	178	TSISISLLDKHSQIINKFYNSVINTLKSTV	SSLLQKEICPLRIFITHSLDVNLQWVDN	237
db	143	DNTSISLLAKKSNFVNLALDSALYLIKRG	LTLPVRKQLCPVLQIISN---TFHPDEISN	199
QY	238	PQ	239	
db	200	PQ	201	

```

RESULT      4
T52638
exportin_1 [validated] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 16-Feb-2001
C:Accession: T52638
R:Haasen, D.; Koehler, C.; Neuhaus, G.; Merkle, T.

```

Plant J. 201, 695-705, 1999
A: title: Nuclear export of proteins in plants: AtXPO1 is the export receptor for leucine-rich ERF domain proteins
A: reference number: 2261149

A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1075 <HAA>
A:Cross-references: EMBL:Y18469; PIRIN:CA89280.1
A:Experimental source: cultivar Columbia; 3wk-old green vegetative tissue
C:Genetics:
A:Gene: xpol
C:Function:
A:Description: functions as nuclear export receptor for proteins carrying leucine-rich
C:Superfamily: Arabidopsis thaliana exportin1

```

Query Match      9.0%; Score 110.5; DB 2; Length 1075;
Best Local Similarity 23.8%; Pred. No. 4.5;
Matches 68; Conservative 46; Mismatches 107; Indels 65; Gaps 15;

QY 5 KIVLLCG-----VLTCSTESILDNLGNLNSVYDKLIEP-VLHGLELTVDNTLGLLEKLK 58
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 593 KIVGCKRRKRVIVGCFNENFENFESFLGCLATTVDNKKHGLIHSEYFSVGNKIQAFENPQ 659
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

[illegible]

RESULT	5
T31557	
Exportin1 (XPO1) protein - Arabidopsis thaliana (fragment)	
N:Alternate names: protein F3K13_170	
C:Species: Arabidopsis thaliana (mouse-ear cress)	

A:Residues: 1-1769 <RAP>
A:CROSS-references: EMBL:X85021; NID:g728698; PIDN:CAA59385.1; PIDs:g728701
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, February 19 1999

C:Genetics:
C:Map position: 101.
C:Keywords: transmembrane protein

Query Match 8.9%; Score 109; DB 2; Length 1769;
Best Local Similarity 19.2%; Pred. No. 11;
Matches 52; Conservative 48; Mismatches 93; Indels 78; Gaps 9;

QY 19 SLDNIGNDLSNVVDKLEPVLIHGLTGVONTLKLGILEKLKVDLGVLQKSSAWLAQAKAQ 78
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 331 SILTFDKEVPCKDFITSYTRSIARYDRS-----KLNIIISLLKK-----ILERY 378
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 79 EAELKLNNTI--SKLPTNTDI-----FGLKTSNLILDVKAEPIDDDCKGNLSPVTAN 131
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 379 EVMLIITDLIYEULEDKSQVELFEFISINEDLVKLCL-----KSGLT----- 425
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 132 VTVAGPIIQIINKASLDLLTAVTIED--PQHQPVAVLGECASDPTSISLSLLOKHS 189
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 426 -----GELPEIRLTSLFTNAVDNTDVKQLSDPV---ETTKKOTASFQFLDKHS 473
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 190 QLIN-----KFVNSVINTLKTSTVSLLQKEICP 217
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 474 ELINTNVMLETGERYKKVLSLTEAIGKYKASSFLTSPFTLESKRIFLLRVTTISP 533
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 218 LIRIFTSHSDVNVIQQVVONPOHKTOLOTLI 248
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 534 AAPALKLSLNIAKIYINSIEKEVNIPTLV 564
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 7
B89944
hypothetical protein SAL447 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: B89944
M:Kuroda, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, M.; Kuroda, A.; Matsuoka, H.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: AB9759; UID:P21311952; PMID:11418146
A:Accession: B89944
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-825 <KUR>
A:CROSS-references: GB:BAB00018; PID:g13701417; PIDN:BAB42711.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
C:Gene: SA1447

Query Match 8.8%; Score 108; DB 2; Length 825;
Best Local Similarity 24.7%; Pred. No. 4.8;
Matches 72; Conservative 39; Mismatches 94; Indels 86; Gaps 15;

QY 16 TSSESLDNLGNDL-----SNVVDKLEPVHLEGLETVDNTLKGIULEKLVLDGLVLOX 66
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 89 TKEAITSYLSSDLFGVGKKTAQNIIVTLG-----DNAINDILD----DHSVLEK 134
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 67 SSAWLAQAKQAQAEKL-LNNVISKLTPNTDI-FGLKTSNS----- 106
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 135 VSGLSKKKQK-QIAEQISAQNSEKIMIRLHDLGCPKLSMAIFYLGDTLTILDRNPY 193
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 107 -LILDVKAEPIDDDCKGNLSPVTANVTVAGPIQGIIINKASLDLLTAVTIED----P 161
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 194 QLTYDIK-----GIGFNKAQDLARNIGIA---YNDNERLKAAL----LYLEECEIKQG 240
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 162 QTHQPVAVLGECA-----SDPTSISLSLDKHSQILNKFNVSINTLKTSTVSLLQKEI 215
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Db 241 HTVLPINVVIDLTVDVINYODEVIEPEKLEMLQYLNEEKRLIIDNEQVAIPSLYSEI 300
 QY 216 CP---LIRIFHSLEVDVNIQO-----VVDNQ-----HKQLQTLI 248
 Db 301 KSVONLFRIKTHTKNLTETEQSDIQMHIGIEDANOVNVAASOKREALOTA 351
 RESULT 8
 A49943
 C:Species: fructose-bisphosphate aldolase (EC 4.1.2.13) - Staphylococcus carnosus (strain TM300)
 C:Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 15-Oct-1999
 C:Accession: A49943; S33358
 R:Witke, C.; Goetz, F.
 J. Bacteriol. 175, 7495-7499, 1993
 A:Title: Cloning, sequencing, and characterization of the gene encoding the class I fructose-bisphosphate aldolase from *Staphylococcus carnosus*
 A:Reference number: A49943; MUID:94042930
 A:Accession: A49943
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-296 <MIT>
 A:Cross-references: EMBL:X71729; NID:g297873; PIDN:CAA50663.1; PID:g297874
 C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 8.6%; Score 106; DB 2; Length 296;
 Best Local Similarity 25.9%; Pred. No. 1.7;
 Matches 50; Conservative 31; Mismatches 64; Indels 48; Gaps 9;

QY 51 KGILEKLVKVDGLVQKSSAWOLAKAKAQAELNNVSKLLPTNTDIFGLKISNLIJD 110
 Db 94 KGIVPELVKVDKGLAEADGVOLMK-PIPOLDKLLDRA-----NERGIFGTAK-RSNILE 145
 QY 111 VKAEPIDDDGKGLNSFPVYANTVAG--PIIGQIIN-----LKASLD-- 150
 Db 146 NNKEAIE--KVVKQOEVAKEIIAAGLVPIIEPEVINAKDKAEATEANLAEAKAELDNL 203
 QY 151 -----LTAVTITET-----DQTHQPVAVLGECAEDPTSLSLDKHDSOIINKF 195
 Db 204 KDOYVMKLTITPTKNAYSELIEHPQVIRVVALSGGYSRDEAN---KILKONDLIASF 260
 QY 196 VNSVINTLKSTVS 208
 Db 261 SRALVSDLNAGQS 273

RESULT 9
 S17448
 C:Species: probable ligand-binding protein KVA3 - rat
 C:Date: 13-Jan-1995 #sequence_revision 11-Jan-1995 #text_change 05-Nov-1999
 C:Accession: S17448
 R:Dear, T.N.; Boehm, T.; Keverne, E.B.; Rabbits, T.H.
 EMBO J. 10, 2813-2819, 1991
 A:Title: Novel genes for potential ligand-binding proteins in subregions of the olfactory bulb
 A:Reference number: S17447; MUID:92007724
 A:Accession: S17448
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-473 <DEA>
 A:Cross-references: EMBL:X60658; NID:g57733; PIDN:CAA43065.1; PID:g57734

Query Match 8.5%; Score 104.5; DB 2; Length 473;
 Best Local Similarity 21.0%; Pred. No. 4.1;
 Matches 60; Conservative 52; Mismatches 93; Indels 81; Gaps 12;

QY 1 LQWKVLVLCGLVTGTGSESL-----DNLGNDLSNVVDKLEPVVLEHGLETVDTNLKGILEK 56
 Db 9 LLMLGLATPCGLLETVGTGLARIDKDELGRATONSLVG-GPILQNVLTGVTVSVNOGLIGA 67
 QY 57--LKVDGLVQLKSSAWOLAKAKAQAELNNVSKLLPTNTDIFGLKISNLI--LDVK 112

Db 68 GLLGGGGLLSYGGLSLVE-----ELSGLKIEELTLPTVSIK 105
 QY 113 AEPIDDDGKGLNSFPVYANTVAGPIIGQIINLKASLDLLTAVTJETDPTQHPVAVIGE 172
 Db 106 LIP---GVGVLQSLHTKVSJHSGPLVG-LQLAAEVNVSSKVALGMSPR-GTPILILKR 160
 QY 173 CASDPTSLISL-----LLDKHSQIINKFVNSVIN-----201
 Db 161 CNTLGHISLTSGLLPTPIHGIIVFOTLCKVLPGLLCPVYDVSLSVNNELLGATLSILVPLG 220
 QY 202---TLKSTVSSILQKEICPLIRIFIHISLDVNVIOQ-----VVDNQ 239
 Db 221 PIGSVETLATL-----PLISNQYIELDINPIVKSIAIGDVIDFEPK 260
 RESULT 10
 T50073
 C:Species: myosin-like coiled-coil protein sp8 [imported] - fission yeast (Schizosaccharomyces fissionis)
 C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
 C:Accession: T50073
 R:McDougal, R.C.; Rajandream, M.A.; Barrell, R.G.; Seeger, K.; Harris, D.
 submitted to the EMBL data library, December 1999
 A:Reference number: Z25034
 A:Accession: T50073
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1727 <MCD>
 A:Cross-references: EMBL:AL133357; PIDN:CAB62414.1; GSPDB:GN00066; SPDB:SPAC1486.04c
 C:Genetics: cosmid c1486
 A:Gene: SPDB:SPAC1486.04c
 A:Map position: 1

Query Match 8.5%; Score 104.5; DB 2; Length 1727;
 Best Local Similarity 24.3%; Pred. No. 23;
 Matches 63; Conservative 41; Mismatches 94; Indels 61; Gaps 11;

QY 17 SESLLDNLGNLSNVYDKLEPV-----LHGLETVDN-----48
 Db 333 SENSLEQLQEKYDVSSELQVKNKNTSVSAGVGLFSPLAQKLSAVQNPETSFTKVYSD 392
 QY 49--TLKGILEKLVKVDGLVQK--SSAWOLAKQ-----KAOEAELNNVSKLLPTNTDIF 99
 Db 393 NMKIQKVSSKLQDLRLTNKFSFCEQVKQRIPVVYKQORSEIVRNNTYNNFLSES-----448
 QY 100 GKITSNLIIDVKAEPIDDDGKGLNSFP-VYANTVAGPIIGQIINLKASLDLLTAVTIE 158
 Db 449--LETSSNNITKVQAELLSTKMKQFACVTLQITASKTQCSDLSSREVICLMAELDLNLFKSR 507
 QY 159 TDKTHQHPVAVLGECAEDPTSLISLDKHSQIINKFVNSVINTLKSTVSSLIQKEICPL 218
 Db 508 NVPAVTO--VALDEYAGNPSTASETLVKN--ELAN-----FSSIKEAVSKTLE-----551
 QY 219 IRIFIHSL--DVNVIOQV 235
 Db 552 LREKVRALQCDVEIQKQTV 570

RESULT 11
 G64242
 C:Species: cytochrome-c accessory protein (hmwl) homolog MG386 - Mycoplasma genitalium
 C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 07-Dec-1999
 C:Accession: G64242
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, D.; Venter, J.C.
 Science 270, 397-403, 1995
 A:Title: The minimal gene complement of Mycoplasma genitalium
 A:Reference number: A64200; MUID:96026346
 A:Accession: G64242

```

Query Match      8.5%; Score 104; DB 2; Length 2710;
Best Local Similarity 20.3%; Pred. NO. 45;
Matches 56; Conservative 63; Mismatches 109; Indels 48; Gaps 11;

QY  5  KLVLCGVLTGTSELLDNLGNLSNVY-----DKLEPVLHEGLETVONTLK--GILEK  56
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB  807  KTLULDASVSPOTKILNNLKLNIHSSIGDYIYIEKLEPVKNIHHSIDDLIDFNLLN  866
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY  57  LKYDLGVLOK-----SSAWQAKAKOAEKELN--NVISKLPNTNTDFG-----  100
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB  867  VSDELYELKKLNNLDEKYLIFEDISKNNSTYSVRFINKNGESVYVEKEKIFSKYSEH  926
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY  101  -----LKISNLSILDVKAEPIDDCCKGNLSPFVTANVTAGPIIGOIINLKASLDULTAV  155
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB  927  ITRKEISTKNSITDVGNNLND---TOLDHTSOVNTLNAFFIQSLIDYSNNKQVLDNL  983
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY  156  TIETDPTQIPVAVLIGECASDPTISISLSDKKHSIIINKEFVNSVITKSTVSLSLQKEI  215
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB  984  STSVKVLQY---AQLFSTGLNTIYDSIQLVN---LISNAVNDTINVLPITTEGI-----  1031
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY  216  CPLIRIFIHSLDGV-VIOQVVD--NPQHKQLQTLI  248
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB  1032  -PIVSTILDGINLGAARKELLEDHDPPLLKRELEAKV  1066
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

RESULT 13
F90067
biochemical protein SA2399 (imported) - Staphylococcus aureus (strain N315)

```

```

494 DAASKLLNTVEE---TTKDVSGLSHKDRKKAVQDHNAEAQDIFCKNLNSLFNNMEELI 549
127 -----PVTANVTVAGPIQOIINIKAS-LDLLTAVTIETDPOTHQPVAVLGECAOPTSI 180
550 KDGSSKQKAMLEVHKTLFGNLLSSVSYALDTITV-----ALGSLTSPENV 596
181 SLSLDKHSHQIINKF-----VNSVINTILKTVSSLLQKEICPLI----- 219
597 STHV----SQIFNMILKEQSLAAEKTVIQELINVLUKTDLLSLEMILSPTVTVSILKINS 652
220 ---RIFIHL-----DVNVIQVVVDNPQHKTQLOTL 247
653 OLKHIFKTSITVADKIEDOKKELDGFLSILCNLNLHELQENTI 694

```

RESULT 15
 C82759
 zinc proteinase XP0816 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: C82759
 C:Anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: C82759
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-930 <SIM>
 A:Cross-references: GB:AE003921; GB:AE003849; NID:g9105710; PIDN:AAF83626.1; GSPDB:CN
 A:Experimental source: strain 9a5c
 R:Simson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carrier
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Perro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
 chado, Junqueira, M.L.; Kemper, E.L.; Kitaajima, J.P.; Krieger, J.E.; Kuranse, E.E.; La
 Jado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.U.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracco, E.C.; Miyaki, C.
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri,
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Tshakoto, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XP0816

Query Match	8.1%	Score 100;	DB 2;	Length 990.
Best local Similarity	21.3%;	Pred. No. 23;	Mismatches 105;	Indels 50; Gaps 8;
Matches	54;	Conservative 45;		
QY	14	TGTSSELDLNGDLSNVDPK - LEPVLHEGLETVDNLTGILEKLKVDGLGVLOKSSAWQL 72 : :	:	
Db	706	SGTNEAIOQTATDLDFQGRWLRPNDNRVILVTGHTTKSIIIPQLEAAFGDWOAPSTIKS 765 : :	:	
QY	73	AKOQAQAEKLLNNVISKLLPTNTDIFGLXISNLI - DVKAEPIDDGKLNLSFPVTAN 131 :	:	
Db	766	HKO -----ITDVAAQGFRIFLIHRPEAQQSILAGLLAPT KD-----PANLE 809 :	:	
QY	132	VTVAGPIIGIINLKASLDL-----LTAVTIETDPOTHQPVAVLGECADPTSISLSL 184 :	:	
Db	810	INVGNFAEGGTFSRLNMNLREEXHWAYGASSVLPNAQGQPYPVFIAVPQTDKTAESIAE 869 :	:	
QY	185	LDKHSQ--INKFVNVSINTLKSTVSSLKEICEPLTRIFTHSLD-----VVNIQ 233 :	:	
Db	870	IQEAQADVINK-----ELTOEVEDKIQQIIRSLPGSYETSGAVLDAVES 915 :	:	
QY	234	VVDNPQHKTQLQTL 247 :	:	

C.Species: Staphylococcus aureus
C.Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C.Accession: F90067
C.Y: R.Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Chui, L.; Ogud
ma, A.; Mizutani-Uh, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.: Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A.Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A.Reference number: A89758; MUID:21311952; PMID:11418146
A.Accession: F90067
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-296 <KUR>
A:Cross-references: CH:BA000018; PID:nl3702563; PID:BA043704.1; GSPDB:GN00149

```

C:Genetics:
A:Gene: SA2399

      Query Match      8.4%:  Score 103.5;  DB 2;  Length 296;
      Best Local Similarity 25.5%:  Pred. No. 2.6;
      Matches 53;  Conservative 30;  Mismatches 68;  Indels 57;  Gaps

QY      41  EGLETVDNTL-KGILEKLVKDVGVIGVIGKSSAMOLAKOKAQAELNNVISKLIPTN-TDI 98
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      83  EKGVTADYLDADKGVGFVKVDKGLAEQNGVOLMKP-----IDNLSLLDRANERHI 134

QY      99  RCLKTSNSLILDVKAEPIDDGKGLNLSRPTVANTVAG--PIIGQIIN----- 144
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      135  FGTAK-RSNILELNEQGIKD--VVEQFEVAKQIIKGLVPIEPEVNIINAKDKAEIKV 191

QY      145  ----LKASLDLLTA-----VTIETDPDTHQPVAVIGECASOPTSISL----- 182
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      192  LKAELKGLDLSNADQLVMLKLTIPTEPNLYK-----ELAEHPNVVRVVVLSGGYSREK 245

QY      183  --SILDKHSQIINKFNYSVINTILKSTVS 208
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      246  ANELLKDNDELIAFSRALASDLRADQS 273

RESULT 14
G02157
kinesin-like spindle protein HKSP - human
C:Species: Homo sapiens (man)
C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C:Accession: G02157
R:Whithead, C.
submitted to the EMBL data library, September 1995
A:Reference number: H00839

```

```

A:Status: preliminary; translated from GenBank/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1056 <WHI>
A:Cross-references: EMBL:U37426; NID:gll71152; PIDN:AAA96132.1; PID:gll71152
C:Genetics:
A:Gene: GDB:KNSL1; Eg5; KSP
A:Map position: GDB:I32856; OMIN:I48760
A:Map position: 10q24.1-10q24.1
C:Superfamily: kinesin-related protein Eg5; kinesin motor domain homology
C:Keywords: ATP; nucleotide binding; P-loop
F:19-36;/Domain: kinesin motor domain homology <KMOT>
F:105-112/Region: nucleotide-binding motif A (P-loop)

      8.2%; Score 100.5; DB 1; Length 1056;
Query Match Best Local Similarity 23.0%; Pred. No. 23;
Matches 65; Conservative 39; Mismatches 89; Indels 89; Gaps

QY    21 LDNLGNDLSNVVDKLPEVLVHGLETVDTNTLKGILBKLKVDGLVLOK--SSAWOLAKOKAQ 78
      || ||| || :||
Db     447 LQCKDSQNKITQEEL-----TTQKHLETKQL-LVKREYITSALESTFKLH 493
QY    79 E-AEKLNNVSKLLPTNTDIFGI--KINSKSLILODKAEPIDD--GKGLNLSF----- 126

```

Db 916 IYRPERDNYIOTL 929

Search completed: August 6, 2002, 17:05:12
Job time: 914 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 17:07:31 ; Search time 32.88 seconds
(without alignments)
292.045 Million cell updates/sec

Title: US-10-020-139-2_COPY_2_249
Perfect score: 1228
Sequence: 1 LQWLKLVLLCGVLTGTSLSL.....NVIQVVNDPNQKTLQTLI 248

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	359.5	29.3	235	1 PSP_MOUSE	P07743 mus musculus
2	192.5	15.7	278	1 PLUN_MOUSE	P97361 mus musculus
3	183.5	14.9	256	1 PLUN_HUMAN	O9np55 homo sapien
4	109	8.9	1769	1 YJK9_YEAST	P42945 saccharomyc
5	106	8.6	295	1 ALF_STACA	O07159 staphylococ
6	104.5	8.5	1727	1 ALM1_SCHPO	Q9uuk5 schizosacch
7	104	8.5	1616	1 P200_MYCGE	Q49429 mycoplasma
8	104	8.5	2710	1 TOXA_CLODI	P16154 clostridium
9	100	8.1	1057	1 EK5_HUMAN	P52732 homo sapien
10	99	8.1	679	1 DNLJ_HAEIN	P43813 haemophilus
11	98.5	8.0	1616	1 RHPO_TOMK2	P89676 tomato mosa
12	98.5	8.0	1616	1 RHPO_TOML	P03587 tomato mosa
13	98.5	8.0	1616	1 RHPO_TOMS1	Q9ykd6 tomato mosa
14	97	7.9	757	1 DNL1_YEAST	P54861 saccharomyc
15	97	7.9	958	1 YGK7_YEAST	P53076 saccharomyc
16	96	7.8	1005	1 RA50_METJA	O58718 methanococc
17	95	7.7	1531	1 YQ38_CAEEL	O09459 caenorhabdi
18	94.5	7.7	1729	1 RRP5_YEAST	O05022 saccharomyc
19	94	7.7	462	1 NIEK_METMP	P71527 methanococc
20	94	7.7	481	1 LBP_HUMAN	P18428 homo sapien
21	94	7.7	490	1 ILVC_BUCAI	P57655 buchnera ap
22	93.5	7.6	868	1 N180_YEAST	P33420 saccharomyc
23	93.5	7.6	1038	1 YKD3_YEAST	P36097 saccharomyc
24	92.5	7.5	529	1 VGLF_SV5	P04849 simian viru
25	92.5	7.5	668	1 BPBC_BAGSU	P42971 bacillus su
26	92.5	7.5	1616	1 RHPO_TOMK1	O9qlt8 tomato mosa
27	92	7.5	1409	1 HAP1_HAEIN	P44596 haemophilus
28	92	7.5	1957	1 YD86_SCHPO	Q10411 schizosacch
29	92	7.5	2329	1 YS89_CAEEL	Q09624 caenorhabdi
30	91.5	7.5	338	1 RLAO_METTL	O52705 methanococc
31	91.5	7.5	4967	1 RYR2_HUMAN	O92736 homo sapien
32	91	7.4	368	1 TSPG_LISMO	P58668 listeria mo
33	91	7.4	2376	1 YIM9_YEAST	P40468 saccharomyc

34	90.5	7.4	519	1 ECTO_RAT	P16573 rattus norv
35	90.5	7.4	761	1 METE_AQUAE	O67606 aquifex ao
36	90.5	7.4	1783	1 Y468_MYCGE	O49460 mycoplasma
37	90	7.3	425	1 NG79_SCHPO	O09793 schizosacch
38	90	7.3	903	1 YB56_METJA	O58556 methanococc
39	90	7.3	1015	1 ITA4_DROME	O9v7a4 drosophila
40	89.5	7.3	559	1 T CPA_YEAST	P12612 saccharomyc
41	89.5	7.3	615	1 YBM2_SCHPO	O10339 schizosacch
42	89	7.2	481	1 LBP_RAT	O63313 rattus norv
43	89	7.2	1526	1 MYS2_SCHPO	O9us16 schizosacch
44	89	7.2	3433	1 UTR0_HUMAN	P46939 homo sapien
45	88.5	7.2	901	1 PIP_DACLA	P49022 lactococcus

ALIGNMENTS

RESULT	J
PSP_MOUSE	
ID	PSP_MOUSE
AC	P07743; STANDARD; PRT; 235 AA.
DT	01-AUG-1988 (Rel. 08, Created)
DT	01-AUG-1988 (Rel. 08, Last sequence update)
DT	01-NOV-1997 (Rel. 35, Last annotation update)
DE	Parotid secretory protein precursor (PSP).
GN	PSP.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID-10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Parotid gland;
RX	MEDLINE-85215456; PubMed-2582349;
RA	Madsen H.O., Hjorth J.P.;
RT	"Molecular cloning of mouse PSP mRNA.";
RL	Nucleic Acids Res. 13:1-13(1985).
RN	[2]
RP	SEQUENCE OF 1-87 FROM N.A.
RC	STRAIN=C3H; TISSUE=Spleen;
RX	MEDLINE-87004556; PubMed-2428613;
RA	Poulsen K., Jakobsen B.K., Mikkelsen B.M.;
RT	Nielsen J.T., Hjorth J.P.;
RT	"Coordination of murine parotid secretory protein and salivary amylase expression.";
RL	EMBO J. 5:1891-1896(1986).
CC	ITS FUNCTION IS NOT KNOWN; HOWEVER, ITS PRODUCTION IS COORDINATED WITH THAT OF SALIVARY AMYLASE.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC	use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: X01697; CA25846.1; -
DR	EMBL: M26807; AAA40009.1; -
DR	EMBL: M26806; AAA40009.1; JOINED.
DR	PIR: A23031; SQMS.
DR	MGI: G97787; Psp.
KW	Parotid gland; Signal.
FT	SIGNAL 1 20
FT	CHAIN 21 235
FT	POTENTIAL.
FT	PAROTID SECRETORY PROTEIN.
FT	SEQUENCE 235 AA; 24753 MW; 23311BAEIE6E2EF3 CRC64;

Query Match 29.3%; Score 359.5; DB 1; Length 235;
Best Local Similarity 33.6%; Pred. No. 4.7e-20;
Matches 80; Conservative 62; Mismatches 79; Indels 17; Gaps 3;

QY 2 QLVKLVLCVLTGTSSESLDNLGNSVVDKLEPVLHEGLTVDNTLKGILEKLVKDL 61
DB 3 QGSLVVLGCLLIGNSESLGELGAVNN-----LKLINPSEAVPQNINLDV 50
QY 62 GVQKSSAMQAKAKAEKLNINNVISKLLPTNTDIFG--LKISNLSLIDVKAEPIDGG 119
DB 51 ELQQAQTSNPLAKNSILET---LNTADLGNLKSFTSLNGLLKLINLKVLDFOAKLSNG 107
QY 120 KGLNLFPPYANTVAGPIIGQILNKLKSLDILLTAVTETDPTQHPVAVLGEACASDPTS 179
DB 108 NGIDLTVPLAGEASLVLPFTGKVDIVSVSLDILNLSIKTNAQTGLPEVTIGKCSNSTD 167
QY 180 ISLSLIDKHSQIINKFNVSINVLKSTVSLLOKEICPLIRIFHSLSLVNVDVQVVDN 237
DB 168 ISISLGRRLPIINSIDGVSTLTSTLTVLQNFCLPQVVLSTINPVSVLQGLLSN 225

RESULT 2
PLUN_MOUSE STANDARD: PRT: 278 AA.
AC P97361;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protein Plunc precursor (Palate lung and nasal epithelium clone
DE protein).
GN PLUNC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=21290678; PubMed=11396972;
RA Leclair E.E., Nguyen L., Bingle L., MacGowan A., Singleton V.,
Ward S.J., Bingle C.D.;
RT "Genomic organization of the mouse plunc gene and expression in the
RT developing airways and thymus";
RL Biochem. Biophys. Res. Commun. 284:792-797(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Palate;
RX MEDLINE=99240770; PubMed=10224143;
RA Weston W.M., Leclair E.E., Trzyna W., McHugh K.M., Nugent P.,
Lafferty C.M., Ma L., Tuan R.S., Greene R.M.;
RA "Differential display identification of plunc, a novel gene expressed
RT in embryonic palate, nasal epithelium, and adult lung";
RL J. Biol. Chem. 274:13698-13703(1999).
RN [3]
RP ERRATUM.
RA Weston W.M., Leclair E.E., Trzyna W., McHugh K.M., Nugent P.,
Lafferty C.M., Ma L., Tuan R.S., Greene R.M.;
RL J. Biol. Chem. 275:8262-8262(2000).
CC -!- FUNCTION: May be involved in the airway inflammatory response
CC after exposure to irritants. May be associated with tumor
CC progression (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- TISSUE SPECIFICITY: Upper airways, nasopharyngeal regions and
CC thymus.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF356785; RAK63069.1;
CC DRL: U69172; AAB63256.1;
CC DR MGD: MGI:1338036; Plunc.
CC Signal; Repeat.

FT SIGNAL 1 19
FT CHAIN 20 278
FT DOMAIN 23 52
FT REPEAT 23 28
FT REPEAT 23 36
FT REPEAT 30 36
FT REPEAT 39 44
FT REPEAT 47 52
FT REPEAT 47 52
FT CARBOHYD 182 282
FT CARBOHYD 228 168
SQ SEQUENCE 278 AA; 28611 MW; 7F40BED9859188FB CRC64;
POTENTIAL.
PROTEIN PLUNC.
4 X 6 AA REPEATS OF G-[LPQ]-[PL]-L-P-L.
REPEAT 1.
REPEAT 2.
REPEAT 3.
REPEAT 4.
N-LINKED (GLCNAC...) (POTENTIAL).
N-LINKED (GLCNAC...) (POTENTIAL).
Query Match 15.7%; Score 192.5; DB 1; Length 278;
Best Local Similarity 23.7%; Pred. No. 1.6e-07;
Matches 67; Conservative 48; Mismatches 107; Indels 61; Gaps 9;
QY 6 LVLCGLVLTGTSSESLD-----NUGNDL-----S 29
DB 7 LVLCGLLAHSTAQLAGLPLPLOGGPPPLNQGPPPLNQGOLLPLAOGPLAVSPALDS 66
QY 30 NVVD---KLEPVLHEGLTVDNTLKGILEKLVKQSSAMQAKAKAEKLN 85
DB 67 NPTDLLAGKFTDLSGGL-----LSGGL-----LGIENIPLLDVIKSGGNSNGLVG 114
QY 86 NVISKL---LPTNTDIFGLKISNLSLIDVKAEPIDGKGLNLSFPVTANTVAGPIIGOI 142
DB 115 GLGKLTSSVPLNLLNLDIKITDPQLLEGLVQSPDGHRLYVTIPLGLTLNVMNPVGS 174
QY 143 INLKASDLLTAVTIETDPO--THOPVAVLGEACASDPTSLSLIDKHSOIINKFNVS 200
DB 175 LQAVKLNITAEVLAVKDNQGRH---LVLDCTHSPGSLIKISLNGVTP-VQSFVDNLT 230
QY 201 NTLKSTVSSLLQKEICPLIRIFHSLSLVNVDVQVVDNPHQKTQ 243
DB 231 GILTKVLPQLIQGVKVCPLVNGILSGLDVTLVHNIAELLHGLQ 273
RESULT 3
PLUN_HUMAN STANDARD: PRT: 256 AA.
AC Q9NP55; Q9NZT0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protein Plunc precursor (Palate lung and nasal epithelium clone
DE protein) (Lung specific X protein) (Nasopharyngeal carcinoma-related
DE protein) (Tracheal epithelium enriched protein).
GN PLUNC OR LUNX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472055; PubMed=11018263;
RA Bingle C.D., Bingle L.;
RT "Characterization of the human plunc gene, a gene product with an
RT upper airways and nasopharyngeal restricted expression pattern";
RL Biochim. Biophys. Acta 1493:363-367(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=21150203; PubMed=11251963;
RA Iwao K., Watanabe T., Fujiwara Y., Takami K., Kodama K.,
Higashiyama M., Yokouchi H., Ozaki K., Monden M., Tanigami A.;
RT "Isolation of a novel human lung-specific gene, LUNX, a potential
RT molecular marker for detection of micrometastasis in non-small-cell
RT lung cancer";
RL Int. J. Cancer 91:433-437(2001).
RN [3]
RP SEQUENCE FROM N.A.
RA He Z.W., Yao K.T., Xu L.G., Ren C.P., Lan K., Xie L., Zhang L.,
Liu W.D., Zhou W., Wang L.;

Cloning a new gene related to nasopharyngeal carcinoma.;
Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

[4]
SEQUENCE FROM N.A.
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.F., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grahnam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levasiaho M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConnachie L.J., McClay K., McMurray A.A.,
RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann M., Sycamore N., Taylor K., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Beck S.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RA *The DNA sequence and comparative analysis of human chromosome 20.*;
RA Nature 414:865-871(2001).

[5]
SEQUENCE FROM N.A.
RA TISSUE=Skeletal muscle;
RA Strausberg R.;
RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

[6]
SEQUENCE OF 95-152; 157-167 AND 214-232.
RA MEDLINE=21317946; PubMed=11425234;
RA Lindahl M., Stahlhof B., Tagesson C.;
RA "Identification of a new potential airway irritation marker, palate
RA lung nasal epithelial clone protein, in human nasal lavage fluid with
RA two-dimensional electrophoresis and matrix-assisted laser
RA desorption/ionization-time of flight.";
RA Electrophoresis 22:1795-1800(2001)
RA -!- FUNCTION: May be involved in the airway inflammatory response
RA after exposure to irritants. May be associated with tumor
RA progression.
RA -!- SUBCELLULAR LOCATION: Secreted.
RA -!- TISSUE SPECIFICITY: Upper airways and nasopharyngeal regions,
RA including trachea and nasal epithelium. Expressed in lung cancers
RA and some other types of cancer.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC EMBL; AF214562; AAG13653.1; -
CC DR EMBL; AF172993; AAF70860.1; -
CC DR EMBL; AB024937; BAA93633.1; -
CC DR EMBL; AF158745; AAF82622.1; -
CC DR EMBL; AL121901; CAC03549.1; -
CC DR EMBL; BC012545; AAH12549.1; -
CC KW Signal. 1 19 POTENTIAL.
CC FT CHAIN 20 256 PROTEIN PLUNC.
CC FT CONFLICT 220 220 Q -> K (IN REF. 1; AAF70860).
CC SEQUENCE 256 AA: 26712 MW: 26712 MW: ED152FBC35315BC CRC64;

Query Match 14.9%; Score 183.5; DB 1; Length 256;
Best Local Similarity 23.8%; Pred. No. 6.9e-07;
Matches 56; Conservative 46; Mismatches 100; Indels 33; Gaps 6;

QY 14 TGTSESLLDNLGNDLSNVVOKLEPVLHGLETVDTNTLKGILEKLVGLVLOKSSAWOLA 73
DB 45 TGLAGSLTNALNSGL-----LSGGL-----LGILENLPLLDIL 77
QY 74 KOKAOEAKELNLNISK---LLPTNTDIFGLKISNLIIDVKAEPIDDDGKGLNLSFPVTA 130
DB 78 KPGGTSGLGLGLGLGKVTSPVIGLNIIIDIKVTDPPOLLEGLVQSPDGHRLVYTIPLGI 137
QY 131 NVTVAGPIIG-QILNKASLDLTAVTIETDPTQTHQPVAVLGECASTPSISLIDKHS 189
DB 138 KLOVNTPLVAGSLRLAVLKLOITAEILAVRDQKRIHL-VLGDCTHSPGSLQISLDGLG 196
QY 190 QI-INKFVSVNTLKTSSVSSLLQKEICPLRIFIHSLDVAVIQVVDNPOHKQ 243
DB 197 PLPIQGLDLSLTGILNKVLPELVQGVNCPVNEVLRGLDITLVHDVNNMLHGLQ 251

RESULT 4

YJK9_YEAST STANDARD; PRT: 1769 AA.
AC P42945;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 200.0 kba protein in GZF3-IME2 intergenic region.
GN YJL109C OR J0808.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=96090136; PubMed=7483851;
RA Rasmussen S.W.;
RA "A 37.5 kb region of yeast chromosome X includes the SWE1, MEF2, GSH1
RA and CSD3 genes, a TCP-1-related gene, an open reading frame similar
RA to the DAL80 gene, and a tRNA(Arg).";
RL Yeast 11-873-883(1995).
CC -!- SIMILARITY: BELONGS TO THE RAP28 FAMILY.
CC -!- SIMILARITY: CONTAINS 1 HEAT REPEAT.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC EMBL; X85021; CAA59385.1; -
CC DR EMBL; 249384; CAA89404.1; -
CC DR SGD; S0003645; YJL109C.
CC DR InterPro; IPR000357; HEAT_repeat.
CC DR PROSITE; PS50077; HEAT_REPEAT; 1.
CC KW Hypothetical protein.
CC REPEAT 1729 1767 HEAT.
CC FT SEQUENCE 1769 AA: 200080 MW: 064480D1D249B241 CRC64;

Query Match 8.9%; Score 109; DB 1; Length 1769;
Best Local Similarity 19.2%; Pred. No. 3;
Matches 52; Conservative 48; Mismatches 93; Indels 78; Gaps 9;

QY 19 SLDNLGNDLSNVVOKLEPVLHGLETVDTNTLKGILEKLVGLVLOKSSAWOLAKOKAQ 78
DB 331 SILTFLDKEDKPKVCDKFTSYSTRARYDRS-----KLNITLSLKK-----IRLRY 378

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39720; AAC71613.1; -
DR EMBL: U02245; AAA03400.1; -
DR EMBL: U02245; AAA03401.1; ALT_FRAME.
DR EMBL: U02175; AAD12458.1; -
DR EMBL: U02126; AAD12402.1; -
DR TIGR: MG386; -
DR Cytadherence; Structural protein; Repeat; Complete proteome.
KW DOMAIN 1205 1389 2 X 32 AA REPEAT.
FT REPEAT 1205 1389 1-1.
FT REPEAT 1205 1389 1-2.
FT DOMAIN 891 1389 2 X 26 AA REPEAT.
FT REPEAT 1161 1186 2-1.
FT REPEAT 1310 1339 2-2.
FT CONFLICT 256 256 P -> S (IN REF. 2).
FT CONFLICT 304 304 S -> F (IN REF. 2).
SQ SEQUENCE 1616 AA; 185678 MW; 6AF76A13AC49E4FF CRC64;

Query Match 8.5%; Score 104.5; DB 1; Length 1727;
Best Local Similarity 24.3%; Pred. No. 6.3;
Matches 63; Conservative 41; Mismatches 94; Indels 61; Gaps 11;

QY 17 SESLLDNLGNLSNVVQKLEPV-----LHGLETVDN----- 48
DB 333 SENSLRELQEKYUSVSELOVAVKENNTSVAGVLESPLAQKLSAVQNPESFTKVVSD 392
QY 49 --TLKGILEKLVQGLVQK--SSAWQLAKQ-----KAQAEKLLNNVSKLPTNTDIF 99
DB 393 NWKLOQKVSSLLKLDRLTNKESFCEQVKQIPVVKQQRSEIVRNIVNMFLES----- 448
QY 100 GLKISNLSILVDKAEPIDDGKGLNLSF-PVTANVTYAGPIIGIOLINKASLDLLTAVTIE 158
DB 449 -LETSNNLTKVQAEILLSTRKQEAQYLQTSRQCSLUSREVLCMAELDHLNETKSR 507
QY 159 TDPOTHOPVAVLGCASDPTSLSLSDKHSQIINKFVNSVINTLKSTVSSLLQKEICPL 218
DB 508 NVPATVQ--VALDEYAQNPTASTETLYNK--ELAN-----FSSIREAVSKTLE----- 551
QY 219 ITRIFIHSL--DVNVIQOVV 235
DB 552 LREKVRALCEDVEIQKQTV 570

RESULT 7
P200_MYCGE STANDARD; PRT; 1616 AA.
AC Q49429; Q49259; Q49298; Q49352; Q49353;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protein P200.
GN MG386.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Doughterty B.A., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bitt K.F., Hu P.-C., Lucher T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 256-427; 432-543 AND 1083-1140 FROM N.A.
RX STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bitt K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -!- FUNCTION: COULD BE A ACCESSORY STRUCTURAL COMPONENT IN
CC CYTADHERENCE (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

Query Match 8.5%; Score 104; DB 1; Length 1616;
Best Local Similarity 20.6%; Pred. No. 6.3; Mismatches 120; Gaps 14;
Matches 63; Conservative 48

QY 17 SESLLDNLGNLSNVVQKLEPVLHGLETVDNLTGLKLVQGLVQKSSAWQLAKQ 76
DB 461 SKEIKDSAKADLSNLSDDISVWKEFGSFTQETQKSVEEKSQVDEILDANDNF---INE 517
QY 77 AQAEKLLNNVIS-----KLLPTNTD----- 97
DB 518 SLFRDEVVNNISQINETSVEQFEPTYSVNEFQEPSEPVVSEKIKETNSDSVNTDL 577
QY 98 --IFGLKISNLSILVDKAEPIDDGKGLNLSFPVTANVTVA-----GPIIGI--IN---- 144
DB 578 TALFSEKLVNELL--TNEYVD-----VNAPFSTETEVKVSSELKPKSELVDEITFINDPK 631
QY 145 ----LKASLDLL-----TAVTIETDPQHPQVAVL-----GECASDPTSI 180
DB 632 PQGLEKYKDFLETEPKSLFDEKTTIVVESEPPFIQDLSLELSDVNDVSKLETKTTSV 691
QY 181 SLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLIRIFIRHSLDNNVIQVVDNPOH 240
DB 692 ELN----HEEIGNEFIN-----LDVSE-KEYEQEP-- 716
QY 241 KTOIQOT 246
DB 717 TTQLET 722

RESULT 8
TOXA_CLODI STANDARD; PRT; 2710 AA.
AC P16154;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Toxin A.
GN TOXA OR TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VPI 10463; PubMed=2109310;
RX MEDLINE=90221894; PubMed=2109310;
RA Sauerborn M., von Eichel-Streiber C.;
RT "Nucleotide sequence of Clostridium difficile toxin A.";

or send an email to license@isb-sib.ch.

```

CC CC      Query Match      8.1%; Score 100; DB 1; Length 1057;
CC DR      Best Local Similarity 24.4%; Pred. No. 7.2;
CC DR      Matches 59; Conservative 34; Mismatches 79; Indels 70; Gaps 13;
CC DR      TIGR: H11100;
CC DR      InterPro: IPR001357; BRCT;
CC DR      InterPro: IPR001679; DNA_Ligase_N;
CC DR      InterPro: IPR004150; DNA_Ligase_OB;
CC DR      InterPro: IPR004149; DNA_Ligase_ZBD;
CC DR      InterPro: IPR004445; HHH;
CC DR      InterPro: IPR003583; HHH;
CC DR      InterPro: IPR003583; HHH;
CC DR      Pfam: PF00533; BRCT; 1;
CC DR      Pfam: PF01653; DNA_Ligase_N; 1;
CC DR      Pfam: PF03120; DNA_Ligase_OB; 1;
CC DR      Pfam: PF03119; DNA_Ligase_ZBD; 1;
CC DR      Pfam: PF00633; HHH; 1;
CC DR      ProDom: PD003944; DNA_Ligase_N; 1;
CC DR      SMART: SM00292; BRCT; 1;
CC DR      SMART: SM00278; HHH1; 2;
CC DR      SMART: SM00532; LIGASE; 1;
CC DR      PROSITE: PS0172; HKCT; 1;
CC DR      PROSITE: PS01055; DNA_LIGASE_N1; 1;
CC DR      PROSITE: PS01056; DNA_LIGASE_N2; 1;
CC DR      PROSITE: PS01056; DNA replication; NAD; Complete proteome.
CC KW      Ligase; DNA repair; DNA replication; NAD; Complete proteome.
CC FT      DOMAIN 601 BRCT.
CC FT      BINDING 125 AMP (BY SIMILARITY).
CC SQ      SEQUENCE 679 AA; 75188 MW; 4C0C33E25D70FF52 CRC64;

RESULT 10
DNLJ HAEIN STANDARD: PRT; 679 AA.
ID DNLJ HAEIN
AC P43813.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA ligase (EC 6.5.1.2) (Polydeoxyribonucleotide synthase [NAD+]).
GN LIGN OR LIG OR H11100.
OS Haemophilus influenzae.
OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Fleischmann A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Kierulff A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitchum W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley G., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -!- FUNCTION: THIS PROTEIN CATALYZES THE FORMATION OF PHOSPHODIESTER
CC LINKAGES BETWEEN 5'-PHOSPHORYL AND 3'-HYDROXYL GROUPS IN DOUBLE-
CC STRANDED DNA USING NAD AS A COENZYME AND AS THE ENERGY SOURCE FOR
CC THE REACTION. IT IS ESSENTIAL FOR DNA REPLICATION AND REPAIR OF
CC DAMAGED DNA (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: NAD(+) + (deoxyribonucleotide)(N) +
CC [deoxyribonucleotide](M) = AMP + nicotinamide nucleotide +
CC -!- SIMILARITY: BELONGS TO THE NAD-DEPENDENT DNA LIGASE FAMILY.
CC -!- SIMILARITY: CONTAINS 1 BRCT DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC CC      Query Match      8.1%; Score 99; DB 1; Length 679;
CC DR      Best Local Similarity 23.1%; Pred. No. 4.8;
CC DR      Matches 54; Conservative 41; Mismatches 71; Indels 68; Gaps 11;
CC QY      23 NLGNDLSNVVDKLEPVHLEGLTVDNTLKGLEKLVGLVLOKSSAWOLA-KKQAEAE 81
CC DB      289 SLGYDIDGTGVLKINDI-----ALQNELGFTSKAPRAIAIYKPPAQBEL 331
CC QY      82 KLLNNV-----ISKLLPTNTDIFGLKISNLSLIDVKAEPIDDDGKGLNLSFPVT 129
CC DB      332 TLLNDVEFQVGRGTAITPVAKLEP-----VFVAGVTVSNATLHG-----DEIERLNAIGDT 384
CC QY      130 ANTVAGPIIGIQLINKASLDLLTAVTIEDTPQHPQVAVLGEACSDPTSSLSLLDKHS 189
CC DB      365 VVTRAGDVIPQII-----GVLHERRPDNAKPIIF-----PTNCPCVC-----DS 423
CC QY      190 QIINKFNVNTLTKSTVSSLOKEICPLIRIFL--HSLDVN-----VIOQVVD 236
CC DB      424 QIIRIEGEAVARCTGGGLFCAARKE---ALKHFVSRKAMDIDGVGKGLIEQLVD 474
CC RESULT 11
CC RPO_TOMK2 STANDARD: PRT; 1616 AA.
CC ID RPO_TOMK2
CC AC P89676; P90349;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE RNA-directed RNA polymerase (EC 2.7.7.48) (183 kDa protein) [Contains:
CC DE Methyltransferase/RNA helicase (MT/HEL) (126 kDa protein)].
CC OS Tomato mosaic virus (strain Kazakh K2) (TOMV) (TMV strain K2).
CC OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
CC OX NCBI_TaxID=138312;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=98012318; PubMed=9454068;
CC RA Belinovich E.V., Genozov E.V., Novikov V.K., Zavrlev S.K.;
CC RT "Properties and structure of the tobacco mosaic virus strain K2
CC genome."
CC RL Mol. Biol. (Mosk) 31:826-830(1997).
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.

```

-!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
BETWEEN CODONS FOR GLN-1116 AND GLN-1118.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: Z92909; CAB07439.1; ;
EMBL: Z92909; CAB07438.1; ;
InterPro: IPR001788; RNA_dep_RNAPol2.
InterPro: IPR002588; V_methyltransf.
InterPro: IPR000606; Viral_helicase.
Pfam: PF00978; RNA_dep_RNAPol2; 1.
Pfam: PF01443; Viral_helicase; 1.
Pfam: PF01660; Vmethyltransf; 1.
Transferrase; RNA-directed RNA polymerase; Helicase; ATP-binding.
CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
METHYLTRANSFERASE/RNA HELICASE.
NP_BIND 833 840 ATP (POTENTIAL).
SEQUENCE 1616 AA; 183614 MW; C2F5CE0C8C965336 CRC64;

Query Match 8.0%; Score 98.5; DB 1; Length 1616;
Best Local Similarity 20.4%; Pred. No. 16;
Matches 40; Conservative 44; Mismatches 79; Indels 33; Gaps 7;

y 56 KLVKDLGVLLQSSAMOLAKQKQAEKLLNNVSKLLPTNTDIFGLKISNLSILDVKAEP 115
b 536 KMSVDMPVLD-----IRKKEETEEMYNALSELVLSKNSDKFDVFSQMCQSLVDVP 588
y 116 IDGK-----GLNLSF--PVTANTVAGPIIGQIINLKASLDLLTAVTIEDPOT 163
b 589 MTAAKVIVAVMSNESGLTLTFEQTPEANVALA-----LQDSEKASDGLVVTSDV 639
y 164 HQPVAVLGECASTPTSTIS--LSLLDKHSQIINKFVNSVINTLSTVSSLOKEICPLIRI 221
y 640 EEP-SIRGSMARGELQAGLSDGVNPESSYTRSEIEISLEQFHMATASSLIHKOMCSI--V 696
y 222 FIHSLDYNVIQVVVDN 237
y 697 YTGPLKVOQMKNFIDS 712

RESULT 12
RRPO_TOML STANDARD; PRT: 1616 AA.
ID RRPO_TOML
AC P03587; O41352;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RNA-directed RNA polymerase (EC 2.7.7.48) (183 kDa protein) [Contains:
Methyltransferase/RNA helicase (MT/HEL) (126 kDa protein)].
OS Tobacco mosaic virus (strain L) (TMV) (TMV strain tomato).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=12252;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85157522; PubMed=6549393;
RA Ohno T., Aoyagi M., Yamanashi Y., Saito H., Ikawa S., Meshi T.,
Okada Y.;
RT "Nucleotide sequence of the tobacco mosaic virus (tomato strain)
genome and comparison with the common strain genome.";
RL J. Biochem. 96:1915-1923(1984).
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CAPPING AND AN RNA HELICASE.
CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
BETWEEN CODONS FOR GLN-1116 AND GLN-1118.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: X02144; CAA26085.1; ;
EMBL: X02144; CAA26082.1; ;
PIR: A04195; WMTM8T.
InterPro: IPR001788; RNA_dep_RNAPol2.
InterPro: IPR002588; V_methyltransf.
InterPro: IPR000606; Viral_helicase.
Pfam: PF00978; RNA_dep_RNAPol2; 1.
Pfam: PF01443; Viral_helicase; 1.
Pfam: PF01660; Vmethyltransf; 1.
Transferrase; RNA-directed RNA polymerase; Helicase; ATP-binding.
CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
METHYLTRANSFERASE/RNA HELICASE.
NP_BIND 833 840 ATP (POTENTIAL).
SEQUENCE 1616 AA; 183564 MW; A8EC8929B5CF7CAF CRC64;

Query Match 8.0%; Score 98.5; DB 1; Length 1616;
Best Local Similarity 20.4%; Pred. No. 16;
Matches 40; Conservative 44; Mismatches 79; Indels 33; Gaps 7;

y 56 KLVKDLGVLLQSSAMOLAKQKQAEKLLNNVSKLLPTNTDIFGLKISNLSILDVKAEP 115
b 536 KMSVDMPVLD-----IRKKEETEEMYNALSELVLSKNSDKFDVFSQMCQSLVDVP 588
y 116 IDGK-----GLNLSF--PVTANTVAGPIIGQIINLKASLDLLTAVTIEDPOT 163
b 589 MTAAKVIVAVMSNESGLTLTFEQTPEANVALA-----LQDSEKASDGLVVTSDV 639
y 164 HQPVAVLGECASTPTSTIS--LSLLDKHSQIINKFVNSVINTLSTVSSLOKEICPLIRI 221
y 640 EEP-SIRGSMARGELQAGLSDGVNPESSYTRSEIEISLEQFHMATASSLIHKOMCSI--V 696
y 222 FIHSLDYNVIQVVVDN 237
y 697 YTGPLKVOQMKNFIDS 712

RESULT 13
RRPO_TOML1 STANDARD; PRT: 1616 AA.
ID RRPO_TOML1
AC Q9WKD6; Q9WJ37;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RNA-directed RNA polymerase (EC 2.7.7.48) (183 kDa protein) [Contains:
Methyltransferase/RNA helicase (MT/HEL) (126 kDa protein)].
OS Tobacco mosaic virus (strain S-1) (TMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=138314;
RN [1]
RP SEQUENCE FROM N.A.
RX Zhou X., Xue C., Chen Q., Qi Y., Li D.;
RT "Complete nucleotide sequence of a Chinese isolate of tomato mosaic
virus.";
RL Submitted: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CAPPING AND AN RNA HELICASE.
CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AJ132845; CAB36997.1; -
DR EMBL: AJ132845; CAB36998.1; -
DR InterPro: IPR001788; RNA_dep_RNAPol2.
DR InterPro: IPR002588; V_methyltransf.
DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
DR Pfam: PF01443; Viral_helicase; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 833 840 ATP (POTENTIAL).
SQ SEQUENCE 1616 AA; 183542 MW; 5DBFB2FADCC5C0C CRC64;

Query Match 8.0%; Score 98.5; DB 1; Length 1616;
Best Local Similarity 20.4%; Pred. No. 16;
Matches 40; Conservative 44; Mismatches 79; Indels 33; Gaps 7;

QY 56 KIKYDGLGVLOKSSANQAKQAEKLLNNVSKLLPTNTDIFGLKTSNLSLIDVKAEP 115
DB 536 KMSVDPVLD-----IRKKMEETEEMYNALSLSKSDKFDVDFVFSOMCSLEVP 588
QY 116 LDDGK-----GLNLSF--PVNTANVTAGPIIGQIINLKASLDLLTAVTETDPQT 163
DB 589 MTAAKVIVAVMSNESGLTTFPEPTAEANVALA-----LQDSEKASDGLVVTSDV 639
QY 164 HOPVAVLGSCASDPTIS--LSLLDKHSQIINKFVNSVINTLKTSSVSSLOKEICPLIRI 221
DB 640 EEP-SIKGSMARGELQAGLSGVDPSSYTRSEETESLEQFHMATASSLIHKQMCST--V 696
QY 222 FTHSLDGVNIQGVNDN 237
DB 697 YTGPLKVQOKNFIDS 712

RESULT 14
DNM1_YEAST STANDARD; PRT; 757 AA.
ID DNM1_YEAST
AC P54861.
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNM1-related protein DNM1 (EC 3.6.1.50).
GN DNM1 OR YLL001W OR L1381.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX K3BI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95348179; PubMed=7622557;
RA Gamble A.E., Kurihara L.J., Vallee R.B., Rose M.D.;
RT "DNM1, a dynamin-related gene, participates in endosomal trafficking
RT in yeast";
RL J. Cell Biol. 130:553-566(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY23;
RX MEDLINE=96405918; PubMed=8810043;
RA Miosga T., Zimmermann F.K.;
RT "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on
RT a 43.7 kb fragment of chromosome XII including an open reading frame
RT homologous to the human cystic fibrosis transmembrane conductance
RT regulator protein CFTR";
RL Yeast 12:693-708(1996).

[3]
RN SEQUENCE FROM N.A.
RP Vandenbol M., Portetelle D., Hilger F.;
RA Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
RL FUNCTION: MICROTUBULE-ASSOCIATED FORCE-PRODUCING PROTEIN THAT
CC PARTICIPATES IN ENOCYTOSIS. DOES NOT APPEAR TO PARTICIPATE IN
CC SECRETION OR VACUOLAR PROTEIN SORTING.
CC -!- CATALYTIC ACTIVITY: GTP + H(2)O = GDP + phosphate.
CC -!- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: L40588; AAA99998.1; -
DR EMBL: X91488; CAA62769.1; -
DR EMBL: 273106; CAA97444.1; -
DR SGD: S0003924; DNM1.
DR InterPro: IPR001401; Dynammin.
DR InterPro: IPR003130; GED.
DR Pfam: PF00350; dynammin; 2.
DR Pfam: PF01031; dynammin_2; 1.
DR Pfam: PF02212; GED; 1.
DR PRINTS: P000195; DYNAMIN.
DR SMART: SM00053; DYNC; 1.
DR SMART: SM00302; GED; 1.
DR PROSITE: PS00410; DYNAMIN; 1.
KW Hydrolase; Motor protein; GTP-binding.
FT NP_BIND 35 42 GTP (POTENTIAL).
FT NP_BIND 175 179 GTP (POTENTIAL).
FT NP_BIND 244 247 GTP (POTENTIAL).
FT CONFLICT 124 124 H -> ISPD (IN REF. 1).
SQ SEQUENCE 757 AA; 84971 MW; EBEF8793C5951770 CRC64;

Query Match 7.9%; Score 97; DB 1; Length 757;
Best Local Similarity 23.1%; Pred. No. 7.8;
Matches 52; Conservative 35; Mismatches 74; Indels 64; Gaps 10;

QY 17 SESL-----LDNLGNDLSNVVDKLEPVLHGLETDNTLKGILEKLYVD-LGVLOKSSAW 70
DB 223 SESLKAREVDPOGKRTIGVTKLD-LMDSGTNLD-ILSGKMYPLKLGFGVGNRS--- 277
QY 71 QIAKQAEAEKLLNNVSKLLPTNTDIF-----GLKISNLSLIDVKAEPID 117
DB 278 ---QDDIQ-----LNKTVEESLDKEEDYFKHPVYRTISTKCGTRYLAKLL----- 320
QY 118 DGKGLNLSFPVTANVTAGPIIGQIINLKASLDLLTAVTETDPOTHPVAVLGECASDP 177
DB 321 -----NQTLISHIRDKLPDIKTKLNTLIS-----QTEQELARYGGVATT 360
QY 178 TSLSLSLLDKHSQIINKFVNSVINTLKTSSVSSLOKEICPLIRIF 222
DB 361 NESRASLV---LQMKNFSTNFISSIDGTSSTINTKELCGGARIY 402

RESULT 15
YGX7_YEAST STANDARD; PRT; 958 AA.
ID YGX7_YEAST
AC P53076;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE DNM1-related protein DNM1 (EC 3.6.1.50).
GN DNM1 OR YLL001W OR L1381.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 17:09:28 ; Search time 111.35 seconds

(without alignments)
385.296 Million cell updates/sec

Title: US-10-020-139-2_COPY_2_249

Perfect score: 1228

Sequence: 1 LQWLKLVLCGVLTGTSESL.....NVIQVVDNPQKTKTQTLI 248

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREXBL_19:

- 1: sp-archaea:
- 2: sp-bacteria:
- 3: sp-fungi:
- 4: sp-human:
- 5: sp-invertebrate:
- 6: sp-mammal:
- 7: sp-mbc:
- 8: sp-organelle:
- 9: sp-phage:
- 10: sp-plant:
- 11: sp-rodent:
- 12: sp-virus:
- 13: sp-vertebrate:
- 14: sp-unclassified:
- 15: sp-virus:
- 16: sp-bacterioph:
- 17: sp-archaeop:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1228	100.0	249	4 Q9BQ00	Q9bq00 homo sapien
2	1220	99.3	249	4 Q96DR5	Q96dr5 homo sapien
3	437	35.6	243	6 P79124	P79124 bos taurus
4	415.5	33.8	240	6 P79125	P79125 bos taurus
5	352.5	28.7	235	11 Q9D734	Q9d734 mus musculus
6	331.5	27.0	235	11 Q63471	Q63471 rattus norv
7	207.5	16.9	206	11 Q63550	Q63550 rattus norv
8	159	12.9	484	4 Q96HK6	Q96hk6 homo sapien
9	150.5	12.3	474	11 Q61114	Q61114 mus musculus
10	148.5	12.1	235	11 Q9D592	Q9d592 mus musculus
11	148.5	12.1	270	11 Q9D794	Q9d794 mus musculus
12	148.5	12.1	270	11 Q9CQX3	Q9cq3 mus musculus
13	145.5	11.8	270	11 Q9D6P0	Q9d6p0 mus musculus
14	132	10.7	232	11 Q9D9J8	Q9d9j8 mus musculus
15	129	10.5	199	4 Q9BQP8	Q9bqp8 homo sapien
16	110.5	9.0	1075	10 Q9SMV6	Q9smv6 arabidopsis

17	108	8.8	825	16	Q9TNO1	Q9tn01 staphylococ
18	104.5	8.5	473	11	Q05701	Q05701 rattus ratt
19	103.5	8.4	296	16	Q99R31	Q99r31 staphylococ
20	100	8.1	666	16	Q9A178	Q9a178 streptococ
21	100	8.1	990	16	Q9PF62	Q9pf62 xyella fas
22	99.5	8.1	718	5	Q71136	Q71136 apis mellif
23	99.5	8.1	1524	1	Q54436	Q54436 staphylothe
24	99	8.1	1441	16	Q9CFL1	Q9cfl1 lactococcus
25	98.5	8.0	842	3	Q9C2C8	Q9c2c8 neurospora
26	98.5	8.0	1116	12	Q991T0	Q991t0 tomato mosa
27	98	8.0	521	11	Q61352	Q61352 mus musculus
28	97.5	7.9	211	17	Q59040	Q59040 methanococ
29	97.5	7.9	387	5	Q9XTU2	Q9xtu2 caenorhabdi
30	97.5	7.9	866	12	Q9WNG5	Q9wng5 tobacco mos
31	97.5	7.9	1116	12	Q9JA03	Q9ja03 tobacco mos
32	97.5	7.9	1616	12	Q9JA04	Q9ja04 tobacco mos
33	96.5	7.9	328	16	Q9KPM2	Q9kpm2 vibrio chol
34	96.5	7.9	2752	5	Q9BJY0	Q9bjy0 plasmodium
35	96	7.8	458	11	Q61351	Q61351 mus musculus
36	96	7.8	624	2	Q49548	Q49548 mycoplasma
37	95.5	7.8	551	12	Q9DUT4	Q9dut4 porcine rub
38	95.5	7.8	2938	11	Q61769	Q61769 mus musculus
39	95	7.7	257	2	Q48995	Q48995 mycoplasma
40	95	7.7	383	5	Q9N2L8	Q9n2l8 caenorhabdi
41	95	7.7	578	16	Q9KS47	Q9ks47 vibrio chol
42	95	7.7	673	16	Q9CKA9	Q9cka9 pasteurella
43	95	7.7	951	10	Q9FVX8	Q9fvx8 oryza sativ
44	95	7.7	6713	16	Q99U54	Q99u54 staphylococ
45	95	7.7	6713	16	Q93IR6	Q93ir6 staphylococ

ALIGNMENTS

RESULT 1
Q9BQ00 PRELIMINARY; PRT: 249 AA.
ID Q9BQ00
AC Q9BQ00
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BA49G10.1 (SIMILAR TO BOVINE SALIVARY PROTEIN BSP30).
GN BA49G10.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
CC NCB_TaxID=9606;
RN 111
RP SEQUENCE FROM N.A.
RA Tracey A.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL121901; CAC03546; J.;
SQ SEQUENCE 249 AA; 27011 MW; E64E0794A1B4DB7D CRC64;

Query Match 100.0%; Score 1228; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 4.3e-82;
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LQWLKLVLCGVLTGTSESLDNLGNLSNVYDKLEPVLHGLETVDTNLTGILEKLYD 60
|||||
DB 2 LQWLKLVLCGVLTGTSESLDNLGNLSNVYDKLEPVLHGLETVDTNLTGILEKLYD 61
|||||
QY 61 LGVLQSSAWLAKQAKQAQAEKLLNNVSKLLPTNTDFFGLKISNLSILDVKAEPIDGK 120
|||||
DB 62 LGVLQSSAWLAKQAKQAQAEKLLNNVSKLLPTNTDFFGLKISNLSILDVKAEPIDGK 121
|||||
QY 121 GLNLSFPVTANVTAGPIIGQIINKASLDLTAVTIEFDPTQHPVAVLGECAADPTSI 180
|||||
DB 122 GLNLSFPVTANVTAGPIIGQIINKASLDLTAVTIEFDPTQHPVAVLGECAADPTSI 181
|||||
QY 181 SLSLDKHSQIINKFNVSNTLTKSVSSLLQKEICPLIRIFIHSLDYNVQVVDNPQH 240
|||||

b 182 SLSLDKHSQINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDVNVVIOQVVDNPOH 241
 y 241 KTQLQTLI 248
 |||||
 b 242 KTQLQTLI 249
 |||||

RESULT 2
 96DR5 PRELIMINARY; PRT; 249 AA.

D Q36DR5;
 C 01-DEC-2001 (TREMUREL. 19, Created)
 T 01-DEC-2001 (TREMUREL. 19, Last sequence update)
 T 01-DEC-2001 (TREMUREL. 19, Last annotation update)
 E PAROTID SECRETORY PROTEIN.
 S Homo sapiens (human)
 C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 C Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 X NCBI_TaxID=9606;
 N [1]
 P SEQUENCE FROM N.A.
 C TISSUE=PAROTID;
 A Venkatesh S.G., Geetha C., Gorr S.U.;
 T "A member of the PSP/plunc family of bPI proteins is expressed in the
 T human parotid gland.";
 L Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 R EMBL: AF432917; AAL28113.1;
 Q SEQUENCE 249 AA; 27110 MW; FD54B624A1A4CA7C CRC64;

Query Match 99.3%; Score 1220; DB 4; Length 249;
 Best Local Similarity 99.6%; Pred. No. 1.6e-81;
 Matches 247; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

y 1 LQWKLVLLCGVLTGTSESLLDNLGNLSNVVDKLEPVVHGLETVDTNLTGILEKLYVD 60
 |||||
 b 2 LQWKLVLLCGVLTGTSESLLDNLGNLSNVVDKLEPVVHGLETVDTNLTGILEKLYVD 61
 |||||

y 61 LGVLOKSSAWOLAKQAKAEKILNNVSKLPTNTDIFGLKISNSLIIDVKAEPIDDDGK 120
 |||||

b 62 LGVLOKSSAWOLAKQAKAEKILNNVSKLPTNTDIFGLKISNSLIIDVKAEPIDDDGK 121
 |||||

y 121 GLNLSFPVTANVTVAGPIGQIINLKASDLTLTAVTETDPQTHQPVAVLGECASTPTSI 180
 |||||

b 122 GLNLSFPVTANVTVAGPIGQIINLKASDLTLTAVTETDPQTHQPVAVLGECASTPTSI 181
 |||||

y 181 SLSLDKHSQINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDVNVVIOQVVDNPOH 240
 |||||

b 182 SLSLDKHSQINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDVNVVIOQVVDNPOH 241
 |||||

y 241 KTQLQTLI 248
 |||||

b 242 KTQLQTLI 249
 |||||

RESULT 3
 979124 PRELIMINARY; PRT; 243 AA.

D Q79124;
 C 01-MAY-1997 (TREMUREL. 03, Created)
 T 01-MAY-1997 (TREMUREL. 03, Last sequence update)
 T 01-DEC-2001 (TREMUREL. 19, Last annotation update)
 DE BSP30.
 JS Bos taurus (Bovine).
 XC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 XC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 XC Bovidae; Bovinae; Bos.
 DX NCBI_TaxID=9913;
 XN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PAROTID GLAND;
 RA Haigh B.J., Wilkins R.J., Wheeler T.T.;
 XA "The cloning and sequencing of two cDNAs coding for alternate forms of

RT BSP30, a bovine member of the Parotid Secretory Protein family.";
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U79413; AAB38282.1;
 SQ SEQUENCE 243 AA; 26877 MW; OC2D8DD45660E11C CRC64;

Query Match 35.6%; Score 437; DB 6; Length 243;
 Best Local Similarity 38.6%; Pred. No. 2.1e-24;
 Matches 95; Conservative 58; Mismatches 85; Indels 8; Gaps 3;

Qy 1 LQWKLVLLCGVLTGTSESLLDNLGNLSNVVDKLEPVVHGLETVDTNLTGILEKLYVD 60
 |||||
 Db 2 VOLWKLVLLCGVLTGTSESLLDNLGNLSNVVDKLEPVVHGLETVDTNLTGILEKLYVD 58
 |||||

Qy 61 LGVLOKSSAWOLAKQAKAEKILNNVSKLPTNTDIFGLKISNSLIIDVKAEPIDDDGK 120
 |||||

Db 59 L-----ESRCLNDVVEETQOTENSLEGLISRIFOVNVRLTGVIRNVQVDPDITFEATSENS 114
 |||||

Qy 121 GLNLSFPVTANVTVAGPIGQIINLKASDLTLTAVTETDPQTHQPVAVLGECASTPTSI 180
 |||||

Db 115 A-DVSIPTADVTVSLPLGELIVKLDNVDLQTSVETDAETGDSRVVVGECPPNPESI 173
 |||||

Qy 181 SLSLDKHSQINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDVNVVIOQVVDNPOH 240
 |||||

Db 174 SLTVLHRRPGLLNDVDFGVNLYRQLVSSVVOHEICPRIRIELLESLDTECIKKLGEPOV 233
 |||||

Qy 241 KTQLQTL 246
 |||||

Db 234 TTOGES 239

RESULT 4
 P79125 PRELIMINARY; PRT; 240 AA.

AC P79125;
 DT 01-MAY-1997 (TREMUREL. 03, Created)
 DT 01-MAY-1997 (TREMUREL. 03, Last sequence update)
 DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)
 DE BSP30.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PAROTID GLAND;
 RA Haigh B.J., Wilkins R.J., Wheeler T.T.;
 RT "The cloning and sequencing of two cDNAs coding for alternate forms of

RT BSP30, a bovine member of the Parotid Secretory Protein family.";
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U79414; AAB38283.1;
 SQ SEQUENCE 240 AA; 26513 MW; 850611DE9E43E358 CRC64;

Query Match 33.8%; Score 415.5; DB 6; Length 240;
 Best Local Similarity 37.9%; Pred. No. 7.8e-23;
 Matches 92; Conservative 56; Mismatches 84; Indels 11; Gaps 5;

Qy 1 LQWKLVLLCGVLTGTSESLLDNLGNLSNVVDKLEPVVHGLETVDTNLTGILEKLYVD 60
 |||||
 Db 2 VOLWKLVLLCGVLTGTSESLLDNLGNLSNVVDKLEPVVHGLETVDTNLTGILEKLYVD 58
 |||||

Qy 61 LGVLOKSSAWOLAKQAKAEKILNNVSKLPTNTDIFGLKISNSLIIDVKAEPIDDDGK 120
 |||||

Db 59 ---LESRCSDVEVEQ--QETENFLEQLISRIFOVNVRLTGVIRNVQVDPDITFEATSENS 113
 |||||

Qy 121 GLNLSFPVTANVTVAGPIGQIINLKASDLTLTAVTETDPQTHQPVAVLGECASTPTSI 180
 |||||

Db 114 A-NVLPIPTADVTVSLPLGELIVKLDNVDLQTSVETDAETGDSRVVVGECPPNPESI 170
 |||||

Qy 181 SLSLDKHSQINKFVNSVINTLKSTVSSLLQKEICPLIRIFIHSLDVNVVIOQVVDNPOH 240
 |||||

Db 171 SLTVLHRSREGLVNDVVDIGVNLARRVSSVVEGELCPHFRELLESIDAECEKVLIGESQD 230
 QY 241 KTO 243
 Db 231 TQO 233

RESULT 5
 Q9D734 PRELIMINARY; PRT: 235 AA.
 ID Q9D734
 AC Q9D734;
 DT 01-JUN-2001 (TREMREL. 17, Created)
 DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
 DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
 DE PAROTID SECRETORY PROTEIN.
 GN Psp.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Sakrali L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK009654; BAB26418.1;
 DR MGD: MGI:97787; Psp
 SQ SEQUENCE 235 AA; 24737 MW; A29D0160268DA0CF CRC64;

Query Match 28.7%; Score 352.5; DB 11; Length 235;
 Best Local Similarity 33.3%; Pred. No. 3e-18;
 Matches 78; Conservative 62; Mismatches 77; Indels 17; Gaps 3;

QY 6 LVVLCGLVGTSESLDNLGNDSNVVDKLEPVLHEGLTVDNLTGKLEKLVLDGLVQ 65
 Db 7 LVVLCGLVGTSESLDNLGNDSNVVDKLEPVLHEGLTVDNLTGKLEKLVLDGLVQ 54
 QY 66 KSSAWQLAKAKAEAEKLNKLVISKLPTNDFG--LKISNLSLIDVKAEPIDGKGLN 123
 Db 55 QATSWPLAKNSILET---LNTADLGNLSFTSLNGLLKLNKLVLDFOAKLSSNGCID 111
 QY 124 LSPVTVANTVAGPIGQIINLKASLDLTAVTIEDTPQTHQPVAVLGECAADPTISLS 183
 Db 112 LTVPLAGEASLVLPFGTKVDIVSVSLDLSINSIKTNAQTGLPEVTIGKCSNTDKTIS 171
 QY 184 LLDKHSQGIINKEFVNSVINTLKSTVSSLLQKEICPLIRIFINSLDNNVIOQVVDN 237
 Db 172 LLGRRLPIINSILDGVSTLTSTLTSTVLGNFLCPLQVYLLSTNPSVQLGLLSN 225

RESULT 6
 Q63471 PRELIMINARY; PRT: 235 AA.
 ID Q63471

AC Q63471;
 DT 01-NOV-1996 (TREMREL. 01, Created)
 DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
 DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
 DE NEONATAL SUBMANDIBULAR GLAND PROTEIN PRECURSOR.
 GN PSP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE-SUBMANDIBULAR GLAND;
 RX MEDLINE=92129360; PubMed=1370829;
 RA Mirels L., Ball W.D.;
 RT "Neonatal rat submandibular gland protein SMG-A and parotid secretory
 RT protein are alternatively regulated members of a salivary protein
 RT multigene family.";
 RL J. Biol. Chem. 267:2679-2687(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE-SUBMANDIBULAR GLAND;
 RX MEDLINE=98129760;
 RA Mirels L., Miranda A.J., Ball W.D.;
 RT "Characterization of the rat salivary-gland B1-immunoreactive
 RT proteins.";
 RL Biochem. J. 330:437-444(1998).
 DR EMBL: M83209; AAC06334.1;
 KW SIGNAL
 FT CHAIN 1 20 POTENTIAL.
 FT CHAIN 21 235 NEONATAL SUBMANDIBULAR GLAND PROTEIN.
 SQ SEQUENCE 235 AA; 24529 MW; 0B36EC779025986E CRC64;

Query Match 27.0%; Score 331.5; DB 11; Length 235;
 Best Local Similarity 31.7%; Pred. No. 1e-16;
 Matches 76; Conservative 62; Mismatches 81; Indels 21; Gaps 5;

QY 2 QLVKLVLCGLVGTSESLDNLGNDSNVVDKLEPVLHEGLTVDNLTGKLEKLVLD 61
 Db 3 QVLSVLCGLVGTSESLDNLGNDSNVVDKLEPVLHEGLTVDNLTGKLEKLVLD 50
 QY 62 GVLSKSAQLAKAKAEAEKLNKLVISKLPTN--TDIFG--LKISNLSLIDVKAEPID 117
 Db 51 GSLSQATTPSAKDSILET--LNKV--ELGNSNGFTPLNGLLKVNFVLDLQGLSS 105
 QY 118 DGKGLNSPPTVANTVAGPIGQIINLKASLDLTAVTIEDTPQTHQPVAVLGECAADP 177
 Db 106 NGKXIDILKPLVFEISFSLPVCPTIDVAVSLDLSNSVQTNATGLPGVTLGKCSNT 165
 QY 178 TSLSLSLLKHSQGIINKEFVNSVINTLKSTVSSLLQKEICPLIRIFINSLDNNVIOQVVDN 237
 Db 166 DKISLSLGRRLPFVNRILDGVSGLLTGAVSILLQNLICPVLQYLLSTNPSVQLGLLSN 225

RESULT 7
 Q63550 PRELIMINARY; PRT: 206 AA.
 ID Q63550
 AC Q63550;
 DT 01-NOV-1996 (TREMREL. 01, Created)
 DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
 DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
 DE NEONATAL SUBMANDIBULAR GLAND PROACINAR CELL PROTEIN PRECURSOR.
 GN SMGB1/SMGB2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE-SUBMANDIBULAR GLAND;
 RX MEDLINE=92129360; PubMed=1370829;
 RA Mirels L., Ball W.D.;

ID	Q9D6P2	PRELIMINARY:	PRT; 235 AA.
AC	Q9D6P2:		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	D2310074B19RIK PROTEIN.		

SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L. M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RA Functional annotation of a full-length mouse cDNA collection.

L Nature 409:685-690(2001).
 R EMBL: AK010115; BAB26710.1; -
 R EMBL: AK009329; BAB26221.1; -
 R EMBL: AK009531; BAB26343.1; -
 R EMBL: AK009562; BAB26360.1; -
 R EMBL: AK009580; BAB26372.1; -
 R EMBL: AK009629; BAB26401.1; -
 R EMBL: AK009695; BAB26444.1; -
 R EMBL: AK009803; BAB26513.1; -
 R EMBL: AK009835; BAB26533.1; -
 R EMBL: AK010012; BAB26542.1; -
 R EMBL: AK010051; BAB26667.1; -
 R MGD: MGI:1914385; 2310021H06RIK.
 Q SEQUENCE 270 AA; 29175 MW; 492A16EEBABA677 CRC64;

Query Match 12.1%; Score 148.5; DB 11; Length 270;
 Best Local Similarity 25.3%; Pred. No. 0.0027;
 Matches 39; Conservative 38; Mismatches 74; Indels 3; Gaps 3;

y 83 LNNVSKLLPTNTDIFGLKISNLSILDVKAEPIDGKGLNLSFPVTANVTAGPIIQOI 142
 b 108 VLGVVSS-IPLLNNILDIRVTPQLLEIGLVQSYDFHRLYVTIPLGFDRVNTLVVGS 166
 y 143 INLKASIDLLTAVTETDPTQHPVAVLGECASDPTISLSLDKHSQIINKFVNSVINT 202
 b 167 LELSVKLDVTAEVYAVRDSYGRSRL-VIGDCIYPPGSLRISLNLRLGLPQN-LIDSLTDI 224
 y 203 LKSTVSSLLQKEICPLIRIFHSLDWNVIQQVVD 236
 b 225 LTRVPLGVQGVVCPVLGVLSLDDVTLAHDVAD 258

RESULT 13
 Q9D6P0 PRELIMINARY; PRT: 270 AA.
 C STRAIN-C57BL/6J; TISSUE-TONGUE;
 X MEDLINE-21085660; PubMed-11217851;
 D Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 T Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 T Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 T Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 E Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 N Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,
 S Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 S Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 C Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 X Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 N Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 N Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 P Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 N Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 C Sesaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 C Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 X Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 N Hayashizaki Y.
 [1]
 "Functional annotation of a full-length mouse cDNA collection.";

C STRAIN-C57BL/6J; TISSUE-TONGUE;
 X MEDLINE-21085660; PubMed-11217851;
 D Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 A Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 A Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 A Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 A Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 A Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,
 A Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 A Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 A Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 A Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 A Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 A Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 A Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 A Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 A Sesaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 A Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 A Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 A Hayashizaki Y.
 [1]
 "Functional annotation of a full-length mouse cDNA collection.";

L Nature 409:685-690(2001).
 R EMBL: AK010115; BAB26730.1; -
 R MGD: MGI:1914385; 2310021H06RIK.
 Q SEQUENCE 270 AA; 29235 MW; A86F02ABBAFBC6 CRC64;

Query Match 11.8%; Score 145.5; DB 11; Length 270;
 Best Local Similarity 25.3%; Pred. No. 0.0045;
 Matches 39; Conservative 38; Mismatches 74; Indels 3; Gaps 3;
 y 83 LNNVSKLLPTNTDIFGLKISNLSILDVKAEPIDGKGLNLSFPVTANVTAGPIIQOI 142
 b 108 VLGVVSS-IPLLNNILDIRVTPQLLEIGLVQSYDFHRLYVTIPLGFDRVNTLVVGS 166
 y 143 INLKASIDLLTAVTETDPTQHPVAVLGECASDPTISLSLDKHSQIINKFVNSVINT 202
 b 167 LELSVKLDVTAEVYAVRDSYGRSRL-VIGDCIYPPGSLRISLNLRLGLPQN-LIDSLTDI 224
 y 203 LKSTVSSLLQKEICPLIRIFHSLDWNVIQQVVD 236
 b 225 LTRVPLGVQGVVCPVLGVLSLDDVTLAHDVAD 258

RESULT 14
 Q9D9J8 PRELIMINARY; PRT: 232 AA.
 C STRAIN-C57BL/6J; TISSUE-TESTIS;
 X MEDLINE-21085660; PubMed-11217851;
 D Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 T Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 T Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 T Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 E Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 N Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,
 S Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 S Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 C Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 X Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 N Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 N Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 P Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 N Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 C Sesaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 C Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 X Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 N Hayashizaki Y.
 [1]
 "Functional annotation of a full-length mouse cDNA collection.";

C STRAIN-C57BL/6J; TISSUE-TESTIS;
 X MEDLINE-21085660; PubMed-11217851;
 D Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 T Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 T Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 T Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 E Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 N Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,
 S Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 S Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 C Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 X Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 N Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 N Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 P Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 N Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 C Sesaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 C Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 X Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 N Hayashizaki Y.
 [1]
 "Functional annotation of a full-length mouse cDNA collection.";

C STRAIN-C57BL/6J; TISSUE-TESTIS;
 X MEDLINE-21085660; PubMed-11217851;
 D Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 T Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 T Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 T Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 E Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 N Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,
 S Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 S Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 C Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 X Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 N Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 N Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 P Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 N Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 C Sesaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 C Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 X Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 N Hayashizaki Y.
 [1]
 "Functional annotation of a full-length mouse cDNA collection.";

L Nature 409:685-690(2001).
 R EMBL: AK006829; BAB24760.1; -
 R MGD: MGI:1920638; 1700058C13RIK.
 Q SEQUENCE 232 AA; 25713 MW; 0D52D24A3076D5DC CRC64;

Query Match 10.7%; Score 132; DB 11; Length 232;
 Best Local Similarity 21.1%; Pred. No. 0.036;
 Matches 55; Conservative 54; Mismatches 88; Indels 54; Gaps 11;

y 3 LK-LVLLCGVLT-----GTSELDNLGNDLSNVVDKLEPVHLEGLETVDNTL 50
 b 4 LWRLLVLLGLLALPSALPQWPGLTKAHKDGSRST-----LARIQAQGLLKLN--A 53
 y 51 KG-----ILEKLKVDLVGLQKSSANQAKQAEKLLNNVSKLLPTNTDIFGLKITS 104
 b 54 EGRIQSMKRLDRNLVSGTVAPGVWGLIGMNFQOQOQISINI-----TNVOL----- 101

QY 105 NSLIIDVKAEPIDGKGLNLSFP---VTANVTAG-----PIICQIINLKASLDLLTAV 155
Db 102 -----DCGGIOMAPPKWFSANITLFEDEFKUPFNSNIITKTHACMGLTAES 148
QY 156 TIETDQTHQPVAVLGEKASDPTSLSLU-LDKHSQITINKEVNSVINTLKSTVSSLLQKE 214
Db 149 WLEKDEFGRRKL-VNCRKMEFSSGCGASMTFETSPKKHFLHNLRFSLGKVIPLNLSVQ 207
QY 215 ICPILRIFIRHSLDQVNIQOVV 235
Db 208 VCPILGEILLRQJLDVKLLKGLV 228

RESULT 15

Q9BOP8 PRELIMINARY: PRT: 199 AA.
AC Q9BOP8:
DT 01-JUN-2001 (TREMHLrel. 17, Created)
DT 01-JUN-2001 (TREMHLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMHLrel. 19, Last annotation update)
DE BA49G10.6 (SIMILAR TO MURINE VON EBNER MINOR SALIVARY GLAND PROTEIN,
DE ISOFORM 1) (FRAGMENT).
GN D1187J4.1
OS Homo sapiens (Human).
OC Fukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tracey A.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMRL; AL121901; CAC03550.1; -;
FT NON_TER 199
SQ SEQUENCE 199 AA: 21517 MW: 49A4CC2143BE04B1 CRC64;

Query Match 10.5%; Score 129; DB 4; Length 199;
Best Local Similarity 24.2%; Pred. No. 0.049;
Matches 54; Conservative 40; Mismatches 89; Indels 40; Gaps 9;
QY 4 WKLVLLCGVLTSTESLNDLNDLSNVVDKLEP--VLHGLETVVDNLTGKILEKLVDL 61
Db 5 WFTLLCGILAAATL-----IQATLSPTAVLILG-----PKVIKEKLTQEL 44
QY 62 -----GVLOKSSAWOLAKQAEKLNINNVISKLPTNTDIEGLKISNLIIDVKAEP 115
Db 45 KDINATSIQQPLLSAMKRPAGGIPVLCGLSVNTVL---KHIIMLKVITANILQQLVKP 101
QY 116 IDGKGLNLSFVTANVTAG---PIICQIINLKASLDLTAVTIFETDQTHQPV-AVLG 171
Db 102 SANDQELLVKIPLD--MVAGNTPLVKTIVEFHMTTE--AQATIRMDTSASGPTRLVLS 156
QY 172 ECASDPTSLSLDKHSGIINKFVNSVINTLKSTVSSLLQKE 214
Db 157 DCATSHGSLRIQLLHKLFLVNALAKQVNNLLVPSLPNLRNQ 199

Search completed: August 6, 2002, 17:09:28
Job time: 930 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

DM protein - protein search, using sw model

Run on: August 6, 2002, 16:52:51 ; Search time 138.55 Seconds

(without alignments)

US-10-020-139-2_COPY_2_249

Perfect score:

Sequence: 1 LQWLKLVLLCGVLTUTSESL.....NVIQVVDPQKRTQLQLI 248

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1228	100.0	249	AAW69221	Human parotid secr
2	1228	100.0	249	AAW60862	Human parotid secr
3	1228	100.0	249	AAE24069	Human PRO1025 prot
4	1228	100.0	249	AAE25765	Human secreted pro
5	1228	100.0	249	AAE75351	Human secreted pro
6	1190	96.9	260	AAE25745	Human protein sequ
7	246	20.0	50	AB341435	Peptide #8941 enco
8	246	20.0	50	AAE62108	Human brain expres
9	246	20.0	50	AAW75111	Human bone marrow
10	246	20.0	50	AAW35227	Peptide #9264 enco
11	189.5	15.4	278	AAE05367	Mouse 28.6 kDa sec

12	183.5	14.9	256	20	AAW06408	Human secreted pro
13	183.5	14.9	256	20	AAW95463	LS170 polypeptide
14	183.5	14.9	256	21	AAW69164	Amino acid sequenc
15	183.5	14.9	256	22	AAU39019	Human secreted pro
16	183.5	14.9	256	22	AAU29210	Human PRO polypept
17	183.5	14.9	256	22	AAW39721	Human polypeptide
18	183.5	14.9	256	22	AAW97366	Human LUNX protein
19	183.5	14.9	264	22	AAW92209	Human digestive sy
20	183.5	14.9	264	22	AAW41507	Human polypeptide
21	168	13.7	187	22	AAW92212	Human digestive sy
22	167	13.6	191	22	AAW92214	Human digestive sy
23	166	13.5	320	22	AAW47214	Human NOV6a protei
24	160	13.0	318	22	AAW47220	Human NOV6b protei
25	160	13.0	484	21	AAW77126	Human neurotransmi
26	160	13.0	565	22	ABG08520	Novel human diagno
27	159	12.9	484	22	AAW63976	Amino acid sequenc
28	158	12.9	484	21	AAW99375	Human PRO1357 (UNQ
29	158	12.9	484	22	AAU29163	Human PRO polypept
30	158	12.9	484	22	AAW87564	Human PRO1357. Ho
31	158	12.9	484	22	AAW66124	Protein of the inv
32	122	9.9	197	21	AAW86219	Human secreted pro
33	122	9.9	221	21	AAW58378	Lung cancer associ
34	108	8.8	751	22	AAW34336	Staphylococcus aur
35	108	8.8	825	22	AAW37370	Staphylococcus aur
36	104	8.5	2710	17	AAW95016	C. difficile toxin
37	104	8.5	2710	19	AAW68387	Clostridium diffic
38	102	8.3	2025	22	AAU34207	Staphylococcus aur
39	102	8.3	3158	22	AAU37018	Staphylococcus aur
40	101	8.2	481	17	AAW16838	Recombinant endoto
41	101	8.2	481	17	AAW16840	Recombinant endoto
42	100	8.1	1057	22	AAW67419	Amino acid sequenc
43	100	8.1	1057	22	AAW47212	Human KSP. Homo s
44	99	8.1	679	22	AAW35585	Haemophilus influe
45	98.5	8.0	1116	15	AAW60608	Tobamovirus replic

ALIGNMENTS

RESULT 1

AAW69221
ID AAW69221 standard; Protein; 249 AA.

XX AC AAW69221;

XX DT 16-OCT-1998 (first entry)

XX XX Human parotid secretory protein.

XX DE Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
KW non-immune defensive disorder; immune system disorder; cancer; human;
KW therapy; diagnosis.

XX OS Homo sapiens.

XX FH Key

XX Peptide Location/Qualifiers

FT Peptide 1..18 /note= "signal peptide"

FT Protein 19..249

FT /note= "mature hPSP"

XX W09828420-A1.

XX 02-JUL-1998.

XX 18-DEC-1997; 97WO-US23522.

XX 23-DEC-1996; 96US-0034429.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Duan R, Ruben SM;

WPI: 1998-377651/32.
N-PSDB: AAV44759.

New nucleic acid encoding human parotid secretory protein or its fragments - useful for diagnosis and treatment of, e.g. digestive and endocrine disorders and for drug screening

Claim 16; Fig 1; 94pp: English.

This sequence is the human parotid secretory protein (hPSP) of the invention. The hPSP DNA is useful for chromosome identification and isolation of the corresponding genomic DNA. The DNA and protein can be used to detect abnormal levels of hPSP (in standard blotting, amplification or immuno assays), particularly for diagnosis of digestive, non-immune defensive, endocrine or immune system disorders. A particular application is diagnosis of cancers of the salivary gland, thymus and pancreas which are associated with high levels of hPSP. The protein is also useful as antifungal, antibacterial, antiparasitic and antiviral agents and may be expressed in vivo from the DNA. The protein, or cells expressing it, are used in screening tests to identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and antisense nucleic acids, which are potentially useful for treating conditions associated with excessive hPSP production. Cells containing the DNA are used to express the recombinant protein and this can be used to raise Ab, useful for diagnosis, therapy, for affinity purification and to identify hPSP-binding proteins.

Sequence 249 AA:

Query Match 100.0%; Score 1228; DB 19; Length 249;
Best Local Similarity 100.0%; Pred. No. 2.3e-100;
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LQWLKLVLCGVLTGTSESLLDNIGNDLSNVVDKLEPVVHLEGLETVDNTLKGILEKLVKD 60
2 LQWLKLVLCGVLTGTSESLLDNIGNDLSNVVDKLEPVVHLEGLETVDNTLKGILEKLVKD 61

61 LGVLQSSAWQLAKQKQAEKLLNNVTSKLLPTNTDIFGLKISNSLTLDVKAPIDDGK 120
62 LGVLQSSAWQLAKQKQAEKLLNNVTSKLLPTNTDIFGLKISNSLTLDVKAPIDDGK 121

121 GLNLSPPTANVTYVAGPIIGQIINLKASLDLLTAVTITDPTQHPVAVLGECDSDPTSI 180
122 GLNLSPPTANVTYVAGPIIGQIINLKASLDLLTAVTITDPTQHPVAVLGECDSDPTSI 181

181 SLSLDKHSQIINKFVNSVINTLKVSTVSSLLQKETCPLIRIFIHSLDYNVITQVVDNPOH 240
182 SLSLDKHSQIINKFVNSVINTLKVSTVSSLLQKETCPLIRIFIHSLDYNVITQVVDNPOH 241

241 KTQLQTLI 248
242 KTQLQTLI 249

RESULT 2
AAW60682
AAW60682 standard; Protein: 249 AA.
AAW60682:
18-SEP-1998 (first entry)
Human parotid secretory protein (hPSP).
Parotid secretory protein; human; cancer; autoimmune disease;
secretory tissue; gastrointestinal tissue; hPSP; Sjorgen's syndrome;
Graves disease; thyroiditis; insulin-dependent diabetes; pancreatitis;
ulcerative colitis; Crohn's disease; atrophic gastritis.
Homo sapiens.
WO9821329-A1.

22-MAY-1998.
07-NOV-1997; 97MO-US20651.
14-NOV-1996; 96US-0749288.
(INCYTE) INCYTE PHARM INC.
Bandman O, Goli SK;
WPI: 1998-297933/26.
N-PSDB: AAV37699.

New parotid secretory protein - useful for, e.g. treatment of cancer and auto-immune disease, particularly of secretory or gastrointestinal tissues

Claim 1; Fig 1A-C; 65pp: English.

This represents a human parotid secretory protein (hPSP). Antagonists that bind specifically to, and modulate activity of hPSP are used to treat cancer and autoimmune diseases particularly of secretory or gastrointestinal tissue, e.g. cancer of salivary gland, thyroid, prostate, breast, gastrointestinal tract of pancreas, Sjorgen's syndrome, Graves disease, thyroiditis, insulin-dependent diabetes, pancreatitis, ulcerative colitis, Crohn's disease and atrophic gastritis. Host cells to produce recombinant hPSP which is used to generate antibodies and to screen for its antagonists. Antibodies are useful directly as antagonists, to transport drugs to hPSP-expressing cells, to detect cells that express hPSP, to monitor patients being treated with hPSP, and for purification of hPSP from natural sources. Expression of hPSP may indicate cell proliferation. hPSP nucleic acid or its fragments are used to detect hPSP-encoding sequences (optionally after amplification by PCR) by hybridisation, particularly for diagnosis and monitoring of disease, but also for mapping the chromosomal sequence.

Sequence 249 AA:

Query Match 100.0%; Score 1228; DB 19; Length 249;
Best Local Similarity 100.0%; Pred. No. 2.3e-100;
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LQWLKLVLCGVLTGTSESLLDNIGNDLSNVVDKLEPVVHLEGLETVDNTLKGILEKLVKD 60
2 LQWLKLVLCGVLTGTSESLLDNIGNDLSNVVDKLEPVVHLEGLETVDNTLKGILEKLVKD 61

61 LGVLQSSAWQLAKQKQAEKLLNNVTSKLLPTNTDIFGLKISNSLTLDVKAPIDDGK 120
62 LGVLQSSAWQLAKQKQAEKLLNNVTSKLLPTNTDIFGLKISNSLTLDVKAPIDDGK 121

121 GLNLSPPTANVTYVAGPIIGQIINLKASLDLLTAVTITDPTQHPVAVLGECDSDPTSI 180
122 GLNLSPPTANVTYVAGPIIGQIINLKASLDLLTAVTITDPTQHPVAVLGECDSDPTSI 181

181 SLSLDKHSQIINKFVNSVINTLKVSTVSSLLQKETCPLIRIFIHSLDYNVITQVVDNPOH 240
182 SLSLDKHSQIINKFVNSVINTLKVSTVSSLLQKETCPLIRIFIHSLDYNVITQVVDNPOH 241

241 KTQLQTLI 248
242 KTQLQTLI 249

RESULT 3
AAB24069
ID AAB24069 standard; Protein: 249 AA.
XX AAB24069;
AC AAB24069;
XX
DT 29-JAN-2001 (first entry)

X Human PRO1025 protein sequence SEQ ID NO:38.
 X Human; tumour; disease; neoplastic disease; neoplastic cell growth;
 X Proliferation; tumorigenesis; identification; cancer; cytostatic;
 X Neurotropic; neuroprotective; antiinflammatory; immunosuppressive;
 X Immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
 X Neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
 X Hypothalamic disorder; glandular disorder; macrophagal disorder;
 X Epithelial disorder; stromal disorder; blastocoele disorder;
 X Inflammatory disorder; immunologic disorder.
 X Homo sapiens.
 X WO2000053755-A2.
 X 14-SEP-2000.
 X 06-JAN-2000; 2000WO-US00376.
 X 08-MAR-1999; 99WO-US05028.
 X 02-JUN-1999; 99WO-US12252.
 X 23-JUN-1999; 99US-0141037.
 X 07-JUL-1999; 99US-0143048.
 X 26-JUL-1999; 99US-0145698.
 X 30-NOV-1999; 99WO-US28313.
 X 20-DEC-1999; 99WO-US30911.
 X 03-JAN-2000; 2000WO-US00219.
 X (GETH) GENENTECH INC.
 X Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;
 X Wetanabe CK, Wood WI;
 X WPI; 2000-572270/53.
 X N-PSDB; AAC58379.
 X Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
 X treatment, diagnosis and prevention of cancer.
 X Claim 61: Fig 26; 286pp; English.
 X The present invention describes an isolated antibody that binds to
 X one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
 X PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
 X PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
 X PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
 X PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
 X growth. The PRO polypeptides and nucleotides are useful in the
 X treatment, diagnosis and prevention of cancer. The antibodies and other
 X anti-tumour compounds may be used to treat various conditions, including
 X those characterised by overexpression and/or activation of the amplified
 X PRO genes. Exemplary conditions or disorders to be treated with such
 X antibodies and other compounds include benign or malignant tumours
 X (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
 X colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic,
 X carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
 X leukaemias and lymphoid malignancies, other disorders such as neuronal,
 X glial, astrocytal, hypothalamic and other glandular, macrophagal,
 X epithelial, stromal and blastocoele disorders, and inflammatory,
 X angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
 X primers and hybridisation probes used in the isolation of the human PRO
 X sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
 X PRO polynucleotide and protein sequences given in the exemplification of
 X the present invention.
 X Sequence 249 AA:

Query Match 100.0%; Score 1228; DB 21; Length 249;
 Best Local Similarity 100.0%; Pred No. 2.3e-100;
 Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQWLKLVIAQVLTGTSESLLDNLGNDLSNWDKLEPVLHGLETVDTNLTGILEKLYD 60
 DB 2 LQWLKLVIAQVLTGTSESLLDNLGNDLSNWDKLEPVLHGLETVDTNLTGILEKLYD 61
 QY 61 LGVLOKSSANOLAKQKAEKLLNNVSKLPTNTDIFGLKISNSLIIDVKAEPIDDK 120
 DB 62 LGVLOKSSANOLAKQKAEKLLNNVSKLPTNTDIFGLKISNSLIIDVKAEPIDDK 121
 QY 121 GLNLSFPVTANVTVAGPTIGOIINLKASLDLLTAVTJETDPQTHQPVAVLGECA 180
 DB 122 GLNLSFPVTANVTVAGPTIGOIINLKASLDLLTAVTJETDPQTHQPVAVLGECA 181
 QY 181 SLSLLDKHSOIINKFVNSVINTLKSTVSSLLQKEICPLIRIFTHSLDQVNVIOQV 240
 DB 182 SLSLLDKHSOIINKFVNSVINTLKSTVSSLLQKEICPLIRIFTHSLDQVNVIOQV 241
 QY 241 KTOLOTLL 248
 DB 242 KTOLOTLL 249
 RESULT 4
 AAB25765
 ID AAB25765 standard; Protein; 249 AA.
 XX AAB25765;
 AC AAB25765;
 XX 28-NOV-2000 (first entry)
 DT Human secreted protein SEQ ID #77.
 XX Human; secreted protein; forensic procedure; gene therapy;
 XX chromosome mapping; cancer; autoimmune disease; cardiovascular disorder;
 XX cystic fibrosis; hypothyroidism; immunological disorder; amyloidosis;
 XX brain disorder; skeletal muscle disorder; eye disorder; obesity;
 XX mitochondrial cytopathy; diabetes; atherosclerosis; Alzheimer's disease;
 XX neurodegenerative disorder; graft rejection; dementia; hyperlipidaemia;
 XX septic shock; impotence.
 XX Homo sapiens.
 OS WO200037491-A2.
 PN 29-JUN-2000.
 PD 20-DEC-1999; 99WO-IB02058.
 PF 22-DEC-1998; 98US-0113686.
 PR 25-JUN-1999; 98US-0141032.
 XX (GEST) GENSET.
 XX Bougueleret L, Dumas J, Duclert A;
 XX WPI: 2000-442637/38.
 XX N-PSDB; AAA87727.
 XX Polynucleotides and polypeptides encoding proteins with signal
 XX peptides, useful in diagnostic, forensic, gene therapy and chromosome
 XX mapping procedures.
 XX Claim 9: Figure 10; 306pp; English.
 XX This sequence represents a human secreted protein amino acid sequence.
 CC The invention relates to sequences AAA87725-A87774 which encode human
 CC secreted proteins AAB25763-B25812. The proteins include signal peptides.
 CC Included in the invention are a host cell containing one of the cDNA
 CC sequences, and a purified antibody capable of binding to one of the
 CC secreted proteins. Also contained in the invention are methods for
 CC storing the sequence data on a computer system, and a method for
 CC identifying features of the cDNA sequences using a computer programme.
 CC The cDNAs are useful for expressing secreted proteins or fragments to
 CC obtain antibodies capable of specifically binding to the secreted

proteins. The cDNAs may also be useful in diagnostic, forensic, gene therapy and chromosome mapping procedures and may be used to design expression vectors and secretion vectors. The proteins of the invention may be used to treat diseases including cancer, autoimmune diseases, cardiovascular disorders, cystic fibrosis, hypothyroidism, immunological disorders, amyloidosis, brain disorders, skeletal muscle disorders, eye disorders, obesity, mitochondrial cytopathies, diabetes, atherosclerosis, neurodegenerative disorders, graft rejection, Alzheimer's disease, dementia, hyperlipidaemia, septic shock and impotence.

Sequence 249 AA;

Query Match 100.0%; Score 1228; DB 21; Length 249;
Best Local Similarity 100.0%; Pred. No. 2.3e-100;
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 LQLWKLVLGCVLTGTSSELDNIGNDLSNVVDKLEPVHLEGLETVDNTLKGILEKLVKD 60
|||||
2 Lqlwklvlcgvltgtseslldnigndlsnvvdklepvlhlegletvntlkigilekikvd 61
|||||
61 LQVLOKSSAWQLAKQKAEKLLNNVTSKLLPTNTDIFGLKISNSLILDVKAEPIDDDGK 120
|||||
62 lqvloksawqlakqkaekllnnvtskllptntdfiglkisnslildvkaepiddgk 121
|||||
121 GUNLSFPVTANVTAGPIIGOIINLKASLDLTLTAVTIEDPQTHQPVAVLGECAADPTSI 180
|||||
122 gnlspfvtanvtavagpiigoiinlkasldltavtiedpqtthqpvavlgecasdpstsi 181
|||||
181 SLSLDKHSQIINKFVNSVINTLKVSTVSSLLQKEICPLIRIFHSLDVNVVQVVDNPOH 240
|||||
182 sllldkhsqinkfvnsvintlksvssllqkeicplirifhslldvnnvqqvvdnpqh 241
|||||
241 KTQLOTLI 248
|||||
242 ktqlqtli 249
|||||

SULT 5

AB75351 standard; protein; 249 AA.

AB75351;

05-APR-2001 (first entry)

Human secreted protein #10.

Secreted protein; prevention; treatment; diagnosis; disease; infection.

Homo sapiens.

WO200100806-A2.

04-JAN-2001.

21-JUN-2000; 2000WO-IB00951.

25-JUN-1999; 99US-0141032.

21-DEC-1999; 99US-0469099.

(GEST) GENSET.

Dumas Milne Edwards J, Bougueleret L, Jobert S;

WPI; 2001-071487/08.

49. Secreted proteins and the nucleic acids encoding them, useful in gene therapy and for detecting similar sequences in samples .

Claim 10; Page 281; 307pp; English.

CC The present invention relates to 49 Secreted proteins and the cDNAs encoding them. The protein and nucleic acids may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate protein expression.

SQ Sequence 249 AA;

Query Match 100.0%; Score 1228; DB 22; Length 249;
Best Local Similarity 100.0%; Pred. No. 2.3e-100;
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQLWKLVLGCVLTGTSSELDNIGNDLSNVVDKLEPVHLEGLETVDNTLKGILEKLVKD 60
|||||
DB 2 Lqlwklvlcgvltgtseslldnigndlsnvvdklepvlhlegletvntlkigilekikvd 61
|||||
OY 61 LQVLOKSSAWQLAKQKAEKLLNNVTSKLLPTNTDIFGLKISNSLILDVKAEPIDDDGK 120
|||||
DB 62 lqvloksawqlakqkaekllnnvtskllptntdfiglkisnslildvkaepiddgk 121
|||||
OY 121 GUNLSFPVTANVTAGPIIGOIINLKASLDLTLTAVTIEDPQTHQPVAVLGECAADPTSI 180
|||||
DB 122 gnlspfvtanvtavagpiigoiinlkasldltavtiedpqtthqpvavlgecasdpstsi 181
|||||
OY 181 SLSLDKHSQIINKFVNSVINTLKVSTVSSLLQKEICPLIRIFHSLDVNVVQVVDNPOH 240
|||||
DB 182 sllldkhsqinkfvnsvintlksvssllqkeicplirifhslldvnnvqqvvdnpqh 241
|||||
OY 241 KTQLOTLI 248
|||||
DB 242 ktqlqtli 249
|||||

RESULT 6

AAM25745

ID AAM25745 standard; Protein; 260 AA.

XX AAM25745;

DT 16-OCT-2001 (first entry)

DE Human protein sequence SEQ ID NO:1260.

KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiagregant; haemostatic; vulnery; antidiabetic; cytotatic;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytotatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.

XX Homo sapiens.

OS WO200153455-A2.

PN 26-JUL-2001.

PF 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HVSE-) HVSEQ INC.

XX

JI Tang YT, Liu C, Drmanac RT;
IX WPI: 2001-457603/49.
JR N-PSDB; AAH95686.

XT Isolated human polynucleotides encoding polypeptides, useful for the
YT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XS Claim 20: Page 260: 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
XC AAM25963. The proteins can have activities based on the tissues and
XC cells they are expressed in, such as: anti-inflammatory; antirheumatic;
XC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
XC central nervous system; viricide; anti-HIV; fungicide; antimutagen;
XC cardiovascular; anhaemic; antiaggregant; haemostatic; vulnerary;
XC antiulcer; osteopathic; dermatologic; antiallergic; antischmatic;
XC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
XC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
XC encoding them can be used in gene therapy, antisense therapy and vaccine
XC production. The proteins and polynucleotides are useful for screening for
XC agonists or antagonists of a protein and for the treatment and diagnosis
XC of disorders associated with the activity of a protein e.g. inflammation,
XC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
XC infections, autoimmunity, genetic diseases, haematopoietic disorders,
XC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
XC osteoporosis, severe combined immunodeficiency, eczema, allergic
XC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
XC Alzheimer's disease, Parkinson's disease, neurodegenerative and
XC neurological disorders.

XS Sequence 260 AA:

Query Match 96.9%; Score 1190; DB 22; Length 260;
Best Local Similarity 97.2%; Pred. No. 5.5e-97;
Matches 241; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

YI 1. LQIMKLVLCVLTGTTSESLLDNLGNDSNVOKLEPLVHEGLTVDNTLKGTLKLYVD 60
YJ 13 lqlwklvllcvgltgtsesllldnlgndslnvdkleplvhegltdvntlkgileklv 72
YK 61 LGVLQKSSAWLAKQKQAEKLLNNVLSKLLPTNTDIFGLKITSNLSLLDVKAEPDGGK 120
YL 73 lgvlgksawqlakqkqaeekllnnvlskllpntdiftgklsnslldvkaepldggk 132
YM 121 GLNLSFPVTANVTAGPTIGQIINLKASLDLLTAVTLETDPQTHQPVAVIGECASDPTS 180
YN 133 glnlsfpvtanvttagplldqilnlrasdlitavtletdpqthhpvagigecardpts 192
YO 181 SLSLLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLRIFIHSLDVNVIOQVDPNQH 240
YP 193 slclldkhsqilnkfvnsvintlkstvssllqkeicplrifihslldvnnvlgqvdnpqh 252
YQ 241 KTLQLOTLL 248
YS 253 ktqlqtlli 260

RESULT 7
ABB41435
D ABB41435 standard; Peptide; 50 AA.

XC ABB41435;
XT 04-FEB-2002 (first entry)
XE Peptide #8941 encoded by human foetal liver single exon probe.
XF Human; foetal liver; gene expression; single exon nucleic acid probe.
XG Homo sapiens.

XX WO200157277-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00669.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR:
XX WPI: 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -

XX Claim 27: SEQ ID NO 34070; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.

XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XS Sequence 50 AA;

Query Match 20.0%; Score 246; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.1e-14;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 PIIGCIINLKASLDLLTAVTLETDPQTHQPVAVIGECASDPTSISLSLLD 186
DB 1 pligciinlkasldlltavtletdpqthhpvagigecadptsisllld 50

RESULT 8
AAM62308
TU AAM62308 standard; Protein; 50 AA.

XX AAM62308;
XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 34413.
XX Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer.

XX Homo sapiens.
XX WO200157275-A2.

XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.
 21-SEP-2000; 2000US-0234687.
 27-SEP-2000; 2000US-0236359.
 04-OCT-2000; 2000GB-0024263.
 (MOLE-) MOLECULAR DYNAMICS INC.
 Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-483446/52.
 Single exon nucleic acid probes for analyzing gene expression in human
 brains -
 Example 4; SEQ ID NO: 34413; 650pp + Sequence Listing; English.
 The present invention provides a number of single exon nucleic acid
 probes which are derived from genomic sequences expressed in the human
 brain. They can be used to measure gene expression in brain cell samples,
 which may enable the diagnosis and improved treatment of nervous system
 diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 epilepsy and cancers. The present sequence is a protein encoded by one of
 the probes of the invention.
 Sequence 50 AA;

Query Match 20.0%; Score 246; DB 22; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.1e-14;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 137 PIIGQIILKASLDLTTAVTETDPTQHPVAVILGECASDPTSLSLD 186
 |||||
 1 piiglinikasldlittavtietdptqhpvavilgecasdptsislld 50

RESULT 9
 AM75111
 AAM75111 standard; Protein; 50 AA.

AAM75111;

06-NOV-2001 (first entry)

Human bone marrow expressed probe encoded protein SEQ ID NO: 35417.

Human; bone marrow expressed exon; gene expression analysis; probe;
 microarray; cancer; leukaemia; lymphoma; myeloma.

Homo sapiens.

WO200157276-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00668.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488900/53.

Human genome-derived single exon nucleic acid probes useful for
 analyzing gene expression in human bone marrow -

XX Example 4; SEQ ID NO: 35417; 658pp + Sequence Listing; English.
 PS
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.
 XX
 SQ Sequence 50 AA;

Query Match 20.0%; Score 246; DB 22; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.1e-14;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 137 PIIGQIILKASLDLTTAVTETDPTQHPVAVILGECASDPTSLSLD 186
 |||||
 1 piiglinikasldlittavtietdptqhpvavilgecasdptsislld 50

RESULT 10
 AAM35227
 ID AAM35227 standard; Protein; 50 AA.

XX AAM35227;

17-OCT-2001 (first entry)

Peptide #9264 encoded by probe for measuring placental gene expression.

Probe; microarray; human; placenta; antenatal diagnosis;
 genetic disorder.

Homo sapiens.

WO200157272-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00663.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488997/53.

Human genome-derived single exon nucleic acid probes useful for
 analyzing gene expression in human placenta -

Claim 27; SEQ ID NO 35496; 654pp; English.

The present invention relates to single exon nucleic acid probes (SENPs;
 see AA131315-AA157546). The present sequence is a peptide encoded by one
 such probe. The probes are useful for producing a microarray for
 predicting, measuring and displaying gene expression in samples derived
 from human placenta. The probes are useful for antenatal diagnosis of
 human genetic disorders.

Sequence 50 AA;

Query Match 20.0%; Score 246; DB 22; Length 50;

Best Local Similarity 100.0%; Pred. NO. 1.1e-14;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 PIIGQIINLKASLDLLTAVTITETDPTQHPVAVLGCASDPTSLSLLD 186
Db 1 pligqinlkasldliltavtietdptqhpvavlgcasdptslslld 50

RESULT 11
AAE05367
ID AAE05367 standard: Protein: 278 AA.
XX AAE05367;
AC AAE05367;
XX
DT 12-SEP-2001 (first entry)
XX
DE Mouse 28.6 kDa secreted protein.
XX
KW Mouse; cytosolic; antiinflammatory; immunoregulatory; tissue integrity;
KW wound healing; immune response; vaccine; cancer; asthma; allergy;
KW cell trafficking; therapy; 28.6 kDa secreted protein.
XX
OS Mus sp.
XX
PN W0200148192-A1.
XX
PD 05-JUL-2001.
XX
PF 21-DEC-2000; 2000MO-N200256.
XX
PR 23-DEC-1999; 99US-0171678.
PR 28-NOV-2000; 2000US-0724864.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Watson JD, Murison JG;
XX
JR WPI: 2001-425665/45.
JR N-PSDB; AAD10135.
XX
PT Novel isolated polypeptide useful to isolate corresponding interacting
PT proteins or other compounds, to quantitatively determine levels of
PT interacting proteins or other compounds, and as therapeutic target
XX
PS Claim 6; Page 92; 101pp; English.
XX
CC The patent discloses novel polynucleotides and their corresponding
CC proteins which play a major role in induction of growth, cell migration
CC and proliferation, cell-cell interaction and the differentiation of
CC tissue-specific cells. These proteins are important in the maintenance
CC of tissue integrity and thus are important in wound healing. They are
CC useful in various assays to determine the biological activity, to raise
CC antibodies, to isolate corresponding interacting proteins or other
CC compounds, to quantitatively determine levels of interacting proteins or
CC other compounds, and as therapeutic target in a whole range of disease
CC states. Compositions comprising the novel proteins of the invention are
CC useful for treating mammalian disorders. Polynucleotides of the invention
CC are useful in genome and physical mapping, in positional cloning of
CC genes, to tag or identify an organism or its reproductive material (as
CC non-disruptive tags for marking organisms), and for the diagnosis and
CC treatment of mammalian diseases which is the consequence of inappropriate
CC expression of kinase genes. They are useful for promoting immune response
CC as part of a vaccine or anti-cancer treatment, as target for cancer
CC treatment, as immunoregulatory and anti-inflammatory molecule, as
CC diagnostic for specific types of cancer and for development of an
CC anti-cancer treatment, and as a target for antagonists in the treatment
CC of diseases such as asthma and allergy. They are also useful to inhibit
CC or enhance the activity of the soluble molecule that binds proteins of
CC the invention, for tissue and neural regeneration, to promote or block
CC cell trafficking, and as anti-inflammatory and/or vaccine adjuvant.
CC The present sequence is 28.6 kDa protein, a secreted protein from mouse.
XX
SQ Sequence 278 AA;

Query Match 15.4%; Score 189.5; DB 22; Length 278;
Best Local Similarity 23.3%; Pred. NO. 1.1e-08;
Matches 66; Conservative 49; Mismatches 107; Indels 61; Gaps 9;

QY 6 LVLLCGVLTGTSESLLD-----NLGNDL-----S 29
Db 7 lllvllcgllahstaqlaglpplgqgppplnqgppplnqgplagglpavspalps 66
QY 30 NVVD---KLEPVLHEGLETVDTNLTGILEKLVLDGLVLOKSSAWLAKAKQAEAKLLN 85
Db 67 nptdlagkitdalsggll-----lsggl-----lgileniplldviksggngslvg 114
QY 86 NVISKL---LPTNTDIFGLKISNSILTDVKAEPIDDKGLNLSFPVTANVTAGPIICQI 142
Db 115 gllgkltssvpllnnildikidpqllelglvqspdgghrlyvtiplgltlnvmpvvgsl 174
QY 143 INLKASLDLLTAVTITETDPTQHPVAVLGCASDPTSLSLLDKHSQIINKFVANSVI 200
Db 175 lqlavklnitaevlavkdnqgrih---lvlgdcthspsgskislingvtp-vqsfldnlt 230
QY 201 NTLKSTVSSILLOKEICPLIRIFHSLDVNVVIOQVVDNPQHKRTQ 243
Db 231 glltkvlpelagkvcpvngilsgldvtlvhniaellihgllq 273

RESULT 12
AAV06408
ID AAV06408 standard: Protein: 256 AA.
XX AC AAV06408;
XX
DT 20-SEP-1999 (first entry)
XX
DE Human secreted protein nh796_1.
XX
KW Secreted protein; nh796_1; human; brain; thalamus; therapy;
KW diagnosis.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 7..19
FT Protein /note= "predicted leader/signal sequence"
FT /note= "mature protein"
PN W09935252-A2.
XX
PD 15-JUL-1999.
XX
PF 31-DEC-1998; 98WO-US27903.
XX
PR 30-DEC-1998; 98US-0222653.
PR 02-JAN-1998; 98US-0070346.
XX
PA (GENY) GENETICS INST INC.
XX
PI Agostino MJ, Collins-Racie LA, Jacobs K, LaVallie ER;
PI McCoy JM, Merberg D, Steininger RJ, Treacy M;
XX
DR WPI: 1999-419350/35.
DR N-PSDB; AAX59356.
XX
PT New polynucleotides encoding secreted human proteins
XX
PS Claim 25; Page 96-97; 100pp; English.
XX
CC This sequence represents a novel human secreted protein, termed
CC nh796_1. The sequence is predicted from a full-length cDNA clone
CC (see AAX59356) isolated from a human adult brain (thalamus) cDNA
CC library. The invention provides cDNA clones (see AAX59352-58)

) 11-OCT-2001.

Search completed: August 6, 2002, 16:52:52
Job time: 334 sec

14	TGTSESLLDNLGNDSNVVDKLEPVVLHGELETVDNTLKGILEKLUVDLGVQLQSSAAQLA	73
QY		
Db		
45	tglsagstna sngl-----lsggl-----lg leen pldlil	77
QY		
Db		
74	KKQAEAEKLLNNVISK--LLPTNTDIFGLKISNLLIDVKAEPIDDGKGLNLSPPVTA	130
QY		
Db		
78	kpgggtsggl lsgl gkvtv p glnnlid kvtcp le lgvsgd nry vt p gi	137
QY		
Db		
131	NVTVAGTIIG-QIINKAKSLDLLTAVTIBTPOTHQPVAVLGEACASPTSIUSLDDKHS	189
QY		
Db		
138	klqvn p vgasll riav kl dtae lav rdk qer ihl-v g d c t s p g s l g l d g l g	196
QY		
Db		
190	QI-INKEVNSVINTLSTKSVSSLLQKEICPLRIFTHSLSDVNVVIOQVVDNPOHQTKQ	243
QY		
Db		
197	p p q g l d s t g l n k v p e l v g v n e v g d l v h d v m n h g l g	251
QY		
Db		

Search completed: August 6, 2002, 16:52:52
Job time: 334 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 16:53:51 ; Search time 51.22 seconds
(without alignments)
118.265 Million cell updates/sec

Title: US-10-020-139-2_COPY_2_249

Perfect score: 1228

Sequence: 1 LQLWKLVLGGLTGTSESL.....NVIQVVDNPQKHTQLOTLI 248

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	104	8.5	2710	1	US-08-480-604A-6	Sequence 6, Appli
2	104	8.5	2710	2	US-08-405-496A-6	Sequence 6, Appli
3	104	8.5	2710	4	US-08-915-136-6	Sequence 6, Appli
4	100	8.1	1057	4	US-09-541-782-10	Sequence 10, Appl
5	96	7.8	481	4	US-08-431-517F-2	Sequence 2, Appli
6	95.5	7.8	210	4	US-09-227-357-195	Sequence 195, App
7	94	7.7	481	1	US-08-186-811-2	Sequence 2, Appli
8	94	7.7	481	1	US-08-311-611A-98	Sequence 98, Appl
9	94	7.7	481	1	US-08-372-783-98	Sequence 98, Appl
10	94	7.7	481	1	US-08-372-105-98	Sequence 98, Appl
11	94	7.7	481	1	US-08-306-473A-98	Sequence 98, Appl
12	94	7.7	481	1	US-08-261-660A-4	Sequence 4, Appli
13	94	7.7	481	1	US-08-209-762-98	Sequence 98, Appl
14	94	7.7	481	1	US-08-473-344-98	Sequence 98, Appl
15	94	7.7	481	1	US-08-274-303-6	Sequence 6, Appli
16	94	7.7	481	1	US-08-377-391A-2	Sequence 2, Appli
17	94	7.7	481	2	US-08-485-445A-98	Sequence 98, Appl
18	94	7.7	481	2	US-08-779-400-2	Sequence 2, Appli
19	94	7.7	481	2	US-08-955-660-2	Sequence 2, Appli
20	94	7.7	481	3	US-09-119-263-98	Sequence 98, Appl
21	94	7.7	481	4	US-08-657-162-98	Sequence 98, Appl
22	94	7.7	481	4	US-09-224-480-98	Sequence 98, Appl
23	94	7.7	481	4	US-09-093-539-98	Sequence 98, Appl
24	94	7.7	481	4	US-08-431-517F-11	Sequence 11, Appli
25	94	7.7	481	4	US-09-146-620-2	Sequence 2, Appli
26	94	7.7	481	4	US-09-395-453-2	Sequence 2, Appli
27	94	7.7	481	5	PCT-US94-02465-98	Sequence 98, Appl

28 94 7.7 481 5 PCT-US94-06931-4 Sequence 4, Appli
29 94 7.7 481 5 PCT-US94-07834-6 Sequence 6, Appli
30 94 7.7 481 5 PCT-US95-00498-98 Sequence 98, Appl
31 94 7.7 481 5 PCT-US95-00656-98 Sequence 4, Appli
32 93 7.6 551 2 US-08-700-548-4 Sequence 4, Appli
33 92.5 7.5 746 2 US-08-785-431-4 Sequence 4, Appli
34 92.5 7.5 746 4 US-09-205-048-4 Sequence 2, Appli
35 92.5 7.5 788 2 US-08-785-431-2 Sequence 2, Appli
36 92.5 7.5 788 4 US-09-205-048-2 Sequence 2, Appli
37 92.5 7.5 789 4 US-08-960-780-6 Sequence 6, Appli
38 92.5 7.5 789 4 US-09-073-898-6 Sequence 6, Appli
39 91 7.4 477 1 US-07-847-562-2 Sequence 2, Appli
40 91 7.4 477 1 US-08-240-328-2 Sequence 2, Appli
41 91 7.4 477 2 US-08-990-849-2 Sequence 2, Appli
42 91 7.4 477 6 5245013-3 Patent No. 5245013
43 91 7.4 479 4 US-08-431-517F-17 Sequence 17, Appli
44 91 7.4 481 2 US-08-215-089-9 Sequence 9, Appli
45 91 7.4 481 4 US-08-431-517F-7 Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480.604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

```

: REFERENCE/DOCKET NUMBER: OPHD-01763
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2710 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-480-604A-6

Query Match      8.5%; Score 104; DB 1; Length 2710;
Best Local Similarity 20.3%; Pred. No. 0.42;
Matches 56; Conservative 63; Mismatches 109; Indels 48; Gaps 11;

QY 5 KLVLCCGLVLTGTSLSLNLGNLSNVV-----DKLEPVLHGLETVDTNLK--GILEK 56
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 807 KTLDDASVSPDKFLLNNLKLNISSIGDYIYKLEPKVKNLIHNSIDDLDEFNLLN 866
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 57 LKVDLGVLOK-----SSAWOLAKQKAEAKLLN--NVISKLLPTNTDIFG----- 100
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 867 VSDLEYELKLLNLDKYLISFEDISKNNSTYSVRFINKSNGESVYVETEKEIFSKYSEH 926
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 101 -----LKISNSLILDVKAEPIDDKGKGLNLSFPVTANVTVAGPIGQIINLKASLDLLTAV 155
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 927 ITKEISTIKNSIITDVNGNLLDN---IQLDHTSQVNTLNAARFFIOSLDYSSNKKVDLNDL 983
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 156 TIETDPQHPVAVIGECASDPTSISSLDDKHSQIINKFVNSVINTLKSTVSSLLQKEI 215
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 984 STSVKVVOLY---AQLFSTGLNTIYDSIQLVN-----LISNAVNDTINVLPITTEGI----- 1031
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 216 CPLIRIFIHSLDYN-VIQOVVD--NPQHKTOLOTLI 248
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1032 -PIVSTILDGILNGAAIKELLDHDPDLKKELEAKV 1066
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 2
US-08-405-496A-6
: Sequence 6, Application US/08405496A
: Patent No. 5919665
: GENERAL INFORMATION:
: APPLICANT: WILLIAMS, JAMES A.
: TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
: TITLE OF INVENTION: NEUROTOXIN
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MEDLEN & CARROLL, LLP
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/405.496A
: FILING DATE: 16-MAR-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/329,154
: FILING DATE: 25-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/161,907
: FILING DATE: 02-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/985,321
: FILING DATE: 04-DEC-1992
: PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: US 07/429,791
: FILING DATE: 31-OCT-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: INGOLIA, DIANE E.
: REGISTRATION NUMBER: 40,027
: REFERENCE/DOCKET NUMBER: OPHD-01308
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2710 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-405-496A-6

Query Match      8.5%; Score 104; DB 2; Length 2710;
Best Local Similarity 20.3%; Pred. No. 0.42;
Matches 56; Conservative 63; Mismatches 109; Indels 48; Gaps 11;

QY 5 KLVLCCGLVLTGTSLSLNLGNLSNVV-----DKLEPVLHGLETVDTNLK--GILEK 56
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 807 KTLDDASVSPDKFLLNNLKLNISSIGDYIYKLEPKVKNLIHNSIDDLDEFNLLN 866
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 57 LKVDLGVLOK-----SSAWOLAKQKAEAKLLN--NVISKLLPTNTDIFG----- 100
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 867 VSDLEYELKLLNLDKYLISFEDISKNNSTYSVRFINKSNGESVYVETEKEIFSKYSEH 926
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 101 -----LKISNSLILDVKAEPIDDKGKGLNLSFPVTANVTVAGPIGQIINLKASLDLLTAV 155
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 927 ITKEISTIKNSIITDVNGNLLDN---IQLDHTSQVNTLNAARFFIOSLDYSSNKKVDLNDL 983
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 156 TIETDPQHPVAVIGECASDPTSISSLDDKHSQIINKFVNSVINTLKSTVSSLLQKEI 215
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 984 STSVKVVOLY---AQLFSTGLNTIYDSIQLVN-----LISNAVNDTINVLPITTEGI----- 1031
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 216 CPLIRIFIHSLDYN-VIQOVVD--NPQHKTOLOTLI 248
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1032 -PIVSTILDGILNGAAIKELLDHDPDLKKELEAKV 1066
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 3
US-08-915-136-6
: Sequence 6, Application US/08915136
: Patent No. 6240960
: GENERAL INFORMATION:
: APPLICANT: KINK, JOHN A.
: APPLICANT: THALLEY, BRUCE S.
: APPLICANT: PADHYE, NISHA V.
: APPLICANT: FIRCA, JOSEPH R.
: APPLICANT: STAFFORD, DOUGLAS C.
: TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
: PREVENTION OF C. DIFFICILE DISEASE
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MEDLEN & CARROLL, LLP
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/915.136
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-6

Query Match 8.5%; Score 104; DB 4; Length 2710;
Best Local Similarity 20.3%; Pred. No. 0.42;
Matches 56; Conservative 63; Mismatches 109; Indels 48; Gaps 11;

QY 5 KLVLLCGVLTGTSSELDLNDLNNV-----DKLEPVHLEGVTDVNTLK--GILEX 56
DB 807 KLLDASVPTKFLNNKLNKLGESSGDYIYEKLEPVKNIHNSIDDLIDFNLEN 866
QY 57 LKVDLGVLOK-----SSAWOLAKOQAEKLLN--NVISKLLPTNTDIFG----- 100
DB 867 VSELYELKLNLDKELVLFSEDISKKNSTYSVRFINKSGESVYVETEKEIFSKYSEH 926
QY 101 -----LKISNSILVDKAEPTDDGKGLNSLPVTANVTVAGPITGOINLKASLDLITAV 155
DB 927 ITKEISTKNSIITDVNGLNDN---IQDHTSOVNTLNAAFFIQSLIDYSSNKKDVLNDL 983
QY 156 TETDPQTHOPVAVLGEACSDPTSISSLKDKHSGIINKFVNSVINTLKSTVSSLLAQKEI 215
DB 984 STSVKVLVY---AQLFSTGLNTLYDSIQLVN---LISNAVNDTINVLPITTEIG[----- 1031
QY 216 CPLRIFHSLDVN-VIOQVVD--NPQHKTOQLTI 248
DB 1032 -DIVSTILDGINLGAIAIKELLDHDPPLKKELEAKV 1066

RESULT 4
US-09-541-782-10
Sequence 10; Application US/09541782
Patent No. 6284480
GENERAL INFORMATION:
APPLICANT: Nislow, Corey
APPLICANT: Sakowicz, Roman
APPLICANT: Beraud, Christophe
TITLE OF INVENTION: Antifungal Assay
FILE REFERENCE: 1015
CURRENT APPLICATION NUMBER: US/09/541,782
CURRENT FILING DATE: 2000-04-03
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10

LENGTH: 1057
TYPE: PRT
ORGANISM: H.sapiens
US-09-541-782-10

Query Match 8.1%; Score 100; DB 4; Length 1057;
Best Local Similarity 24.4%; Pred. No. 0.26;
Matches 59; Conservative 34; Mismatches 79; Indels 70; Gaps 13;

QY 21 LDNLGNDLSNVVDKLEPVHLEGVTDVNTLKGLILEKLVLDGLVLOK--SSANQAKAKQAO 78
DB 447 LDCKSLQNKKTUELE-----TQKHLOQTKLOL-VKEEYITSALSTEEXLH 493
QY 79 E-AEKLLNNVISKLLPTNTDIFGL--KISNSILVDKAEPTDD--GKGLNLSF----- 126
DB 494 DAASKLLNTVEE---TTKDVSGLSKLDKRAVDQHNAAQDIFGKNLNSLFNNWEEIJ 549
QY 127 -----PVTANVTVAGPIIGOLINKAS-LDLLTAVTETDPTQHPVAVLGEACSDPTSI 180
DB 550 KUGSSKQKAMLEVHKTLFGNCLSSSVSALDTITTV-----ALGSLTSIPENV 596
QY 181 SLSLDKHSQIINKF-----VNSVINTLKSTVSSLLQKEICPLIRIFIHSLD 227
DB 597 STHV---SQIFNMILKEQSLAAESKTVLQELINVLTLLSSLEWILSPTV---VSILK 649
QY 228 VN 229
DB 650 IN 651

RESULT 5
US-08-431-517F-2
Sequence 2; Application US/08431517F
Patent No. 6265187
GENERAL INFORMATION:
APPLICANT: Scott, Randal W
APPLICANT: Marra, Marian N
TITLE OF INVENTION: RECOMBINANT ENDOTOXIN-NEUTRALIZING PROTEINS
FILE REFERENCE: 1103/11307US01
CURRENT APPLICATION NUMBER: US/08/431,517F
CURRENT FILING DATE: 1995-05-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 481
TYPE: PRT
ORGANISM: human
OTHER INFORMATION: No. 6265187e; human LRP-b DNA and amino acid sequence
US-08-431-517F-2

Query Match 7.8%; Score 96; DB 4; Length 481;
Best Local Similarity 19.6%; Pred. No. 0.21;
Matches 51; Conservative 52; Mismatches 107; Indels 50; Gaps 11;

QY 7 VILCGVLTGTSSELDLNDLNNVVDK-LEPVHLEGVTDVNTLKGI-LEKLVLDGLVL 64
DB 11 ILLALLTSTPEALGANPGL-VARITDKGLQVAAQEGLLALQSELLRITLPDFTGDLRIP 69
QY 65 OKSSAWOLAKOQAEKLLNNSVISKLLPTNTDIFGLKISNSILVDKAEPTDDGKGLNL 124
DB 70 HVCRG-----RYEFHSLNHSCHELLHSALRPV-PGOGLSL 103
QY 125 SPVTVANTVTAG--PIQOIINLKASLDL-LTAVTETD-----PQTHOPVAVLGEACSD 176
DB 104 SIS-DSSIRVQGRKVRKSFKEKLGSDVSVKGISISVNLGSSSESGRPVTVTASSCSD 162
QY 177 PTSISLSDKHSQIINKFVNSVIN---TLKSTVSSLLQKEICPLIRIFIHSLD----- 228
DB 163 IADVEVMSGDFGWLNLNLFHNOIESKFOKYLEICEMIQKSVSSDLQPYIOTLPVTTET 222
QY 229 ----NVIQVVDNPQHKTOQL 244

Db 223 DSFADIDYSILVEAPRATQOM 242

RESULT 6

US-09-227-357-195

Sequence 195, Application US/09227357

Patent No. 6342581

GENERAL INFORMATION:

APPLICANT: Fischer et al.

TITLE OF INVENTION: 123 Human Secreted Proteins

FILE REFERENCE: P2010P1

CURRENT APPLICATION NUMBER: US/09/227,357

CURRENT FILING DATE: 1999-01-08

EARLIER APPLICATION NUMBER: PCT/US98/13684

EARLIER FILING DATE: 1998-07-07

EARLIER APPLICATION NUMBER: 60/051,926

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/052,793

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/051,925

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/051,929

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/052,803

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/052,732

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/051,931

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/051,932

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/051,916

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/051,930

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/051,918

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/051,920

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/052,733

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/052,795

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/051,919

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/051,928

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/055,722

EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,723

EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,948

EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,949

EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,953

EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,950

EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,947

EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,964

EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/056,360

EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,684

EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,984

EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,954

EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/058,785

EARLIER FILING DATE: 1997-09-12

EARLIER APPLICATION NUMBER: 60/058,664

EARLIER FILING DATE: 1997-09-12

EARLIER APPLICATION NUMBER: 60/058,660

EARLIER FILING DATE: 1997-09-12

EARLIER APPLICATION NUMBER: 60/058,661

EARLIER FILING DATE: 1997-09-12

NUMBER OF SEQ ID NOS: 672

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 195

LENGTH: 210

TYPE: PKT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (210)

OTHER INFORMATION: xaa equals stop translation

US-09-227-357-195

Query Match 7.8%; Score 95.5; DB 4; Length 210;

Best Local Similarity 19.2%; Pred. No. 0.069;

Matches 44; Conservative 46; Mismatches 86; Indels 53; Gaps 8;

QY 3 LKLVLLCGVLT-----GTSELLDNLGNDLSNVVDKLEPVLRHGLTVDNLT 51

Db 4 LWRLIFLGLLALPLAPHKQWPGLAQAHDRNKSRLARTIAOGL--IKHNAESRIQNIHF 61

QY 52 GILEKLVLDGLVQLKSSAMQAKQAQAEKLLNNVSKLLPTNTDIFGLKIS----- 104

Db 62 G--DRLNASSAQVAPGLVGLWSGRKHQQOQESSINITNQL----DCGGIQISPHKEWFS 115

QY 105 --NSLILDVKAEPIDGKGLNLSFPVTANVTAGPIGOIINLKASLDLLLTAVTITDPO 162

Db 116 ANISLEFDLELRPSFD-----NNITKMAHMSIVVEFWLEKDEF 154

QY 163 THOPVAVLGEACSDPTSSISLLDKHSQIKNFVNSVINTLKSTVSSLL 211

Db 155 GRDL-VIGKDAEPSSVHVAIL---TEAIPKMQOFLYNLAKENLQKVL 199

RESULT 7

US-08-186-811-2

Sequence 2, Application US/08186811

Patent No. 5484705

GENERAL INFORMATION:

APPLICANT: White, Mark Leslie

APPLICANT: Carroll, Stephen Fitzhugh

APPLICANT: Ma, Jeremy Kam-kuen

TITLE OF INVENTION: METHOD FOR QUANTIFYING LBP IN BODY FLUIDS

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/186,811

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Sharp, Jeffrey S.

REGISTRATION NUMBER: 31,879

REFERENCE/DOCKET NUMBER: 27129/31843

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300


```

; FILING DATE: 11-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/209,762
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/183,222
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 27129/32415
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 481 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: "rLBP"
US-08-372-783-98

```

```

Query Match          7.7%; Score 94; DB 1; Length 481;
Best Local Similarity 19.6%; Pred. No. 0.33;
Matches 51; Conservative 52; Mismatches 107; Indels 50; Gaps 11;

QY 7 VLICGVLGTSTSLDNLGNLSNVVDK-LEPVLHREGLETVDNTLKG1-LEKLVLDLGV 64
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 11 ILLALLTSTPEALGANPGL-VARITDKGLQYAAQEGLLALQSELLRTITLPDFTGDLRIP 69

QY 65 QXSSAWQLAKQAQAEKLLNNVISKLLPTNTDIFGLKISNLSILDVKAEPIDDGKGLN 124
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 70 HVGRG-----RYEFHSLNIHSCHELLSALRPV-PGQGLSL 103

QY 125 SPVTVANTVAG--PIIGQIINIKASLDL-LTAVTIETD-----POTHQPVAVLGECA 176
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 104 SIS-DSSIRVQGWKVRKSEFFKLQSDVSVKGISISVNLGSGESSGRPIVTASSCSD 162

QY 177 PYSISLSDKHSQIINKFVNSVIN----TLKSTVSSILQKEICPLIRIFIHSLDV---- 228
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 163 IADVEVDKSGDGLGMLNLPHNQIESKFQVLESKICEMIQKSVSSDLQPYLQTLPTVTEI 222

QY 229 ----NVIQOVVDNPOHKTQL 244
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 223 DSFADIDYSLVEAPRATAQM 242

```

```

RESULT 10
US-08-372-105-98
; Sequence 98, Application US/08372105
; Patent No. 5627153
; GENERAL INFORMATION:
; APPLICANT: Little, Roger G.
; APPLICANT: Lim, Edward
; APPLICANT: Lambert, Lewis H.
; APPLICANT: Scannon, Patrick J.
; TITLE OF INVENTION: Anti-Fungal Materials and Methods
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/372,105
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/273,540
; FILING DATE: 11-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/209,762
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/183,222
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 27129/32415
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 481 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: "rLBP"
US-08-372-105-98

```

```

Query Match          7.7%; Score 94; DB 1; Length 481;
Best Local Similarity 19.6%; Pred. No. 0.33;
Matches 51; Conservative 52; Mismatches 107; Indels 50; Gaps 11;

QY 7 VLICGVLGTSTSLDNLGNLSNVVDK-LEPVLHREGLETVDNTLKG1-LEKLVLDLGV 64
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 11 ILLALLTSTPEALGANPGL-VARITDKGLQYAAQEGLLALQSELLRTITLPDFTGDLRIP 69

QY 65 QXSSAWQLAKQAQAEKLLNNVISKLLPTNTDIFGLKISNLSILDVKAEPIDDGKGLN 124
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 70 HVGRG-----RYEFHSLNIHSCHELLSALRPV-PGQGLSL 103

QY 125 SPVTVANTVAG--PIIGQIINIKASLDL-LTAVTIETD-----POTHQPVAVLGECA 176
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 104 SIS-DSSIRVQGWKVRKSEFFKLQSDVSVKGISISVNLGSGESSGRPIVTASSCSD 162

QY 177 PYSISLSDKHSQIINKFVNSVIN----TLKSTVSSILQKEICPLIRIFIHSLDV---- 228
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 163 IADVEVDKSGDGLGMLNLPHNQIESKFQVLESKICEMIQKSVSSDLQPYLQTLPTVTEI 222

QY 229 ----NVIQOVVDNPOHKTQL 244
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 223 DSFADIDYSLVEAPRATAQM 242

```

```

RESULT 11
US-08-306-473A-98
; Sequence 98, Application US/08306473A
; Patent No. 5652332
; GENERAL INFORMATION:
; APPLICANT: Little, Roger G.
; TITLE OF INVENTION: Biologically Active Peptides from
; TITLE OF INVENTION: Functional Domains of Bactericidal/
; TITLE OF INVENTION: Permeability-Increasing Protein and
; TITLE OF INVENTION: Uses Thereof
; NUMBER OF SEQUENCES: 226
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: Suite 3000, 10 S. Wacker Drive

```



```

; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/306,473A
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/209,762
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/183,222
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McDonnell, John J.
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 93,1133-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 481 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: "rLBP"
; US-08-306-473A-98

```

```

Query Match          7.7%; Score 94; DB 1; Length 481;
Best Local Similarity 19.6%; Pred. No. 0.33;
Matches 51; Conservative 52; Mismatches 107; Indels 50; Gaps 11;

Qy 7 VLLCGVLTGTSISLDNLGNDLSNVVDK-LEPVLHGLETVDTNLKGI-LEKLKVDLGLV 64
   || || || || || || || || || || || || || || || || || || || ||
Db 11 ILLALLTSTPEALGANPGL-VARITDKGLQYAAQEGLLALQSELLRLTLPDFTGDLRP 69

Qy 65 QKSSAWLAKQAQAEKLLNNVSKLLPTNTDIFGKISNLSILDVKAEPIDDKGLNL 124
   || || || || || || || || || || || || || || || || || || || ||
Db 70 HVGRG-----RYEFHSLNTHSCCELLHSALRPV-PGQGLSL 103

Qy 125 SFPVTANVTAG--PITGQIINLKASLDL-LTAVTIETD-----PQTHOPVAVLGCASD 176
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 104 SIS-DSSIRVQGRWKVKRSFFKLGQSFVSVKGISISVNLGSSSGRPTVTASSCSD 162

Qy 177 PTSLSLLDKHSQIINKFVNSVIN----TLKSTVSSLLQKEICPLIRIFIHSLDY---- 228
   || || || || || || || || || || || || || || || || || || || ||
Db 163 IADVEVMSDGLGWLNLNLFHNIQIESKPKVLESRICEMIQKSVSSDLQPYLOTLPVTTEI 222

Qy 229 ----NVIQQVVDNPOHKTOL 244
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 DSFADIDYSLVEAPRATAQM 242

```

```

RESULT 12
US-08-261-660A-4
; Sequence 4, Application US/08261660A
; Patent No. 5731415
; GENERAL INFORMATION:
; APPLICANT: Gazzano-Santoro, Helene
; APPLICANT: Treofan, Georgia
; APPLICANT: Town, Patrick W.
; TITLE OF INVENTION: Lipopolysaccharide Binding Protein Derivatives
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,660A
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp, Jeffrey S.
; REGISTRATION NUMBER: 31,879
; REFERENCE/DOCKET NUMBER: 27129/32137
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3836
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 481 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: "rLBP"
; US-08-261-660A-4

```

```

Query Match          7.7%; Score 94; DB 1; Length 481;
Best Local Similarity 19.6%; Pred. No. 0.33;
Matches 51; Conservative 52; Mismatches 107; Indels 50; Gaps 11;

Qy 7 VLLCGVLTGTSISLDNLGNDLSNVVDK-LEPVLHGLETVDTNLKGI-LEKLKVDLGLV 64
   || || || || || || || || || || || || || || || || || || || ||
Db 11 ILLALLTSTPEALGANPGL-VARITDKGLQYAAQEGLLALQSELLRLTLPDFTGDLRP 69

Qy 65 QKSSAWLAKQAQAEKLLNNVSKLLPTNTDIFGKISNLSILDVKAEPIDDKGLNL 124
   || || || || || || || || || || || || || || || || || || || ||
Db 70 HVGRG-----RYEFHSLNTHSCCELLHSALRPV-PGQGLSL 103

Qy 125 SFPVTANVTAG--PITGQIINLKASLDL-LTAVTIETD-----PQTHOPVAVLGCASD 176
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 104 SIS-DSSIRVQGRWKVKRSFFKLGQSFVSVKGISISVNLGSSSGRPTVTASSCSD 162

Qy 177 PTSLSLLDKHSQIINKFVNSVIN----TLKSTVSSLLQKEICPLIRIFIHSLDY---- 228
   || || || || || || || || || || || || || || || || || || || ||
Db 163 IADVEVMSDGLGWLNLNLFHNIQIESKPKVLESRICEMIQKSVSSDLQPYLOTLPVTTEI 222

Qy 229 ----NVIQQVVDNPOHKTOL 244
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 DSFADIDYSLVEAPRATAQM 242

```

```

RESULT 13
US-08-209-762-98
; Sequence 98, Application US/08209762
; Patent No. 5733872
; GENERAL INFORMATION:
; APPLICANT: Little, Roger G.
; TITLE OF INVENTION: Biologically Active Peptides from
; TITLE OF INVENTION: Functional Domains of Bactericidal/Permeability-Increasing
; TITLE OF INVENTION: Protein and Uses Thereof
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000

```


GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 17:05:12 ; Search time 66.51 Seconds
(without alignments)
310.618 Million cell updates/sec

Title: US-10-020-139-2_COPY_35_249

Perfect score: 1061

Sequence: 1 KLRPVLHGLETVDMTLKGI.....NVIQVVVDNPOHKTQLTLI 215

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIK_71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	294.5	27.8	235	1 SOMS	parotid secretory
2	263.5	24.8	235	2 B42337	parotid secretory
3	154.5	14.6	206	2 A42337	submandibular gland
4	106	10.0	296	2 A49943	fructose-bisphosph
5	106	10.0	1769	2 S53378	probable membrane
6	103.5	9.8	296	2 F90067	hypothetical prote
7	99.5	9.4	1727	2 T50073	myosin-like coiled
8	98.5	9.3	1615	1 WMT88T	180k protein - tom
9	97.5	9.2	211	2 D64505	hypothetical prote
10	96.5	9.1	1345	2 H90975	hypothetical prote
11	96.5	9.1	2660	2 E85822	probable invasin 2
12	95	9.0	679	2 D64182	DNA ligase (NAD+)
13	95	9.0	1130	2 T19148	hypothetical prote
14	94.5	8.9	825	2 B89944	hypothetical prote
15	94.5	8.9	1729	2 S57596	ribosomal RNA proc
16	93.5	8.8	1038	2 S37854	hypothetical prote
17	93.5	8.8	2938	2 T30249	cell proliferation
18	93	8.8	524	2 PC6003	surface membrane p
19	93	8.8	1237	2 AC1583	interanal protein
20	92.5	8.7	414	2 T27045	hypothetical prote
21	92.5	8.7	958	2 S64249	hypothetical prote
22	92	8.7	2383	2 D64962	probable membrane
23	91.5	8.6	473	2 S17448	probable ligand-bi
24	91	8.6	443	2 B82209	GGDEF family prote
25	91	8.6	454	2 T18765	hypothetical prote
26	91	8.6	578	2 B82204	methyl-accepting c
27	91	8.6	2710	2 A37052	toxin A - Clostrid
28	90.5	8.5	1254	2 E82064	conserved hypotet
29	90.5	8.5	1668	2 A60272	IGA-specific metal

30 30 90 8.5 529 1 VGNZSP
31 89.5 8.4 1056 1 G02157
32 89.5 8.4 2819 2 A90551
33 89 8.4 368 2 A11254
34 89 8.4 481 2 I56246
35 89 8.4 1075 2 T52638
36 89 8.4 1075 2 T51557
37 88.5 8.3 1509 2 B89985
38 87.5 8.2 602 2 E71562
39 87.5 8.2 707 2 T29559
40 87.5 8.2 990 2 C82759
41 87.5 8.2 6713 2 B89921
42 87 8.2 310 2 F90011
43 87 8.2 621 2 T15859
44 86.5 8.2 256 2 S77810
45 86.5 8.2 668 2 I39902

ALIGNMENTS

RESULT 1

SOMS

parotid secretory protein precursor - mouse

N:Alternate names: PSP

C:Species: Mus musculus (house mouse)

C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 22-Jun-1999

C:Accession: A23031, I53236

R:Madson, H.O.; Hargrave, J.P.

Nucleotide: 1-13, 1985

Cloning of mouse PSP mRNA.

031; MUID:85215456

01697; NID:953810; PIDN:CAA5846.1; PID:G758163
B.K.; Mikkelsen, B.M.; Harmark, K.; Nielsen, J.T.; Hjorth,
B6
murine parotid secretory protein and salivary amylase expr-
16; MUID:87004556
translated from GB/EMBL/DBJ

References: GB:M26807; NID:G200556; PIDN:AAA40009.1; PID:9554264

C:Comment: PSP is the most abundant protein in the parotid gland. Its function is n

C:Genetics:

A:Gene: Psp

A:Map position: 2

A:Introns: 41/1

A:Note: list of introns may be incomplete

C:Superfamily: parotid secretory protein

C:Keywords: parotid gland; saliva

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-235/Product: parotid secretory protein #status predicted <MAT>

Query Match

Best Local Similarity 27.8%; Score 294.5; DB 1; Length 235;

Matches 69; Conservative 55; Mismatches 74; Indels 13; Gaps 4;

QY 3 EPLVHGLETVDMTLK-----GILEKLKLVGLQKSSAMOLAKQAQAEKLLNNVI 55
DB 19 ESLLGE-LGSANVLKILNPPSAVPQNLVQLQQTATWPLAKNSILET---LNTAD 74
QY 56 SKLLPTNTDIFG--LKISNLILDKVAEPIDGKGLNLSFPVTVANTVAGPIGINK 113
DB 75 LGNLKSFSLNGLLLKINLKLSSNGNGIDLVPLAGEASLVLPFGTKTVDIS 134
QY 114 ASLDLLTAVTIEDPOTHQPVAVLGECDPFSISLSLDKHSQIINKFVSVINTLKST 173
DB 135 VSLDLSINSIKTNAQTGPEVTIGKSSNTDKISISLGLRLPIINSILDGVTLLTST 194

```

Qy 174 VSSLQKEICPLIKIFTHSLDVNVVQVVDN 204
      :|::| |::| :|: |::| :|: |
Db 195 LSTVLQNFCLPIQYVLSTLNPVLOGLLSN 225

RESULT 2
B42337
Parotid secretory protein precursor - rat
N:Alternate names: PSP
C:Species: Rattus norvegicus (Norway rat)
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 20-Aug-1999
C:Accession: B42337
R:Mirels, L.; Ball, W.D.
J. Biol. Chem. 267. 2679-2687, 1992
A:Title: Neonatal rat submandibular gland protein SMG-A and parotid secretory protein an
A:Reference number: A42337; MUID:92129360
A:Accession: B42337
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-235 <MIK>
A:Cross-references: GB:M83209; NID:q206456; PIDN:AAC06334.1; PID:q206457
A:Note: sequence extracted from NCBI backbone (NCBIN:78709, NCBI:P:78710)
C:Genetics:
A:Gene: Psp
C:Superfamily: parotid secretory protein
C:Keywords: parotid gland; saliva

```

	Query Match	24.8%	Score 263.5	DB 2	Length 235
	Best Local Similarity	29.6%	Pred. No. 8.3e-14		
	Matches 59	Conservative	56	Mismatches 75	Indels 9
					Caps 4
Qy	10	LETVDNTLKGILEKLVKVDLVGLVQKSSAQWLAKQKAEKLLNNVISKLLPTN--TDIFG	67		
		! : : : : : ! : ! : : : ! : ! : : : ! : ! : : : ! : ! : : !			
Db	32	LDLINSPEAAQNLDVGLVQSGQATTWPAKDSILET---LNKV--ELGNSGFTPLNG	86		
		! : : : : ! : ! : : : ! : ! : : : ! : ! : : : ! : ! : : !			
Qy	68	--LKISNLSLDVKAEPIDCGKGNLSFFVTANTVTVAGPIGQIINLKASLDLLTAVTIE	125		
		! : : : : ! : ! : : : ! : ! : : : ! : ! : : : ! : ! : : !			
Db	87	LLLRVNFRLVDLQAGLSSNGKDLDLPLVFEISFSLVPIGPTLDVAVSLDLLNSYSVQ	146		
		! : : : : ! : ! : : : ! : ! : : : ! : ! : : : ! : ! : : !			
Qy	126	TDPTHQFVAVIGECASDPTSIUSLIDKHSQIINKVKNVSNVINTLKTSTVSLQKEICPL	185		
		! : : : : ! : ! : : : ! : ! : : : ! : ! : : : ! : ! : : !			
Db	147	TNAQTGLPGVFTLKGCSGNTDKISISLLCRRLPFYNRIUDGVSGLLTGAVSILLQNI	206		
		! : : : : ! : ! : : : ! : ! : : : ! : ! : : : ! : ! : : !			
Qy	186	IRIFHSLDENVVIQQVVUN	204		
		! : : : : ! : ! : : : ! : ! : : : ! : ! : : : ! : ! : : !			
Db	207	LOYLLSTMSGAIQGLJASN	225		

RESULT 3
A42337 submandibular gland protein A - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 02-Sep-2000
C:Accession: A42337
R:Mirets, B.; Ball, W.D.
J. Biol. Chem. 267, 2679-2687, 1992
A:Title: Neonatal rat submandibular gland protein SMG-A and parotid secretory protein an
A:Reference number: A42337; MUID:92129360
A:Accession: A42337
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-206 <MR>
A:Cross-references: GHR:W83210; NID:g206989; PIDN:AAC12783.1; PID:g206990
A:Note: sequence extracted from NCBI backbone (NCBIN:78707, NCBIp:78708)
C:Superfamily: parotid secretory protein

```
Query Match      14.6%; Score 154.5; DB 2; Length 206;
Best Local Similarity 24.5%; Pred. NO. 3.6e-05;
Matches 34; Conservative 42; Mismatches 60; Indels 3; Gaps 1;
```

Qy 68 LK1SNSLILDVKAEPIDDGKGLNSFPVTANVTVAGPIIGQIINLKASLDLLTAVTIETD 127

D6		: : : : : : :	LKIKOLRLNLNHEVSPNGDEVTLKHPMALNSLUPARDLTDDVSISMEAITSFALEND	125
QY	128	POTHOPVAVLGECADPTSTLSLLDKHSGQINKFNVSVINTLKSTVSSLLQKEICPLIR	187	
D6	126	: : : : : : : : : :	PKTGRRLVMQRCSLNTDNTSISLLNPKSNFYNALDSALYLKRGTLTPYRROLCPVLQ	185
QY	188	JFIHSLDYNVIQQVVDPNQ	206	
D6	186	: : : : : : : :	LIITSN---TFHPDEISNPQ	201

```

RESULT      4
A:Accession: A49943
C:Species: fructose-bisphosphate aldolase (EC 4.1.2.13) - Staphylococcus carnosus (strain TM300)
C:Species: Staphylococcus carnosus
C:Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 15-Oct-1999
C:Accession: A49943; S33358
R:Witke, C.; Goetz, F.
J. Bacteriol. 175, 7495-7499, 1993
A:Title: Cloning, sequencing, and characterization of the gene encoding the class I
A:Reference number: A49943; NUID:94042930
A:Accession: A49943
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-296 <WIT>
A:Cross-references: EMBL:X71729; NID:g297873; PID:g297874
C:Keywords: aldehyde-lyase; carbon-carbon lyase

```

Query Match	10.0%	Score 106:	DB 2:	Length 296:
Best Local Similarity	25.9%	Pred. No.	0.43:	
Matches	50:	Conservative	31:	Mismatches 64: Indels 48: Gaps
Qy	18	KGILEKLVDLGVLYQKSSAWQLAKQAQAEKLLNNVWISKLLPTNTDIFGLKISNLSILD	77	
Db	94	KGIVPELVKDLGLAEAGCVQLMK--PIFDLKLDDRA-----NERGIFGTM--RGNILE	145	
Qy	78	VKAEPIDDDKGLNLSEFPYTVANTVYAG--PIIQGIN-----LKASLD--	117	
Db	146	NNKEAIE--KVYKQQPEVAKETIAGLVPIEPEVINAKDKAEIAENLAEAIAKELDNL	203	
Qy	118	-----LLTAVTIT-----DPQTHQVAVLGCASDPTSI:LSLLDKHSIQINKF	162	
Db	204	KDQYVYMLKLTITPKVNAYSEIIEHPQVIRVVALSGGYSRDEAN---KILKQNDGLIASF	260	
Qy	163	VNSVINTLKSTVS	175	
Db	261	SRAIVSDLNAAQS	273	

```

RESULT      5
S53378
probable membrane protein YJL109c - yeast (Saccharomycetes cerevisiae)
N:Alternate names: hypothetical protein J0808
C:Species: Saccharomycetes cerevisiae
C>Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 21-Jul-2000
C:Accession: S53378; S56887; S57359
R:Rasmussen, S.W.
submitted to the EMBL Data Library, February 1995
A:Description: A 37.5 kb region of yeast chromosome X includes the SME1, ME1,
A:Sequence number: S53376
A:Accession: S53378
A:Molecule type: DNA
A:Residues: 1-1769 <RAS>
A:Cross-references: EMBL:X85021; NID:g728698; PID:g728701
R:Rasmussen, S.W.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56876
A:Accession: S56887
A:Molecule type: DNA
A:Residues: 1-1769 <RAW>
A:Cross-references: EMBL:249384; NID:q1008292; PID:q1008293; MIPS:YJL109C

```

R:Rasmussen, S.W.
Yeast 11, 873-883, 1995
A:Title: A 37.5 kb region of yeast chromosome X includes the SML1, ME2, GSH1 and CSD3 genes
A:Reference number: S57357; MUID:96090136
A:Accession: S57359
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1769 <RAF>
A:Cross-references: EMBL:X85021; NID:q728698; PIDN:CAA59385.1; PID:q728701
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995
C:Genetics:
A:Map position: 10L
C:Keywords: transmembrane protein

Query Match 10.0%; Score 106; DB 2; Length 1769;
Best Local Similarity 19.2%; Pred. No. 4.5;
Matches 45; Conservative 44; Mismatches 73; Indels 72; Gaps 8;

QY 23 KLVKLVGVLOKSSANOLAKQAEKLLNNV--SKLLPTNTOI-----FGLKISNLSLI 75
Db 362 KNIITLSLKK-----IRLRYEVRLLITDIYLSFIEDKSLVFLFYFISINEDLV 415

QY 76 LDVKAEPIDGKGLNLSFPVTANVTAGPIIGQIINKKASLDLLTAVTIFD--POTHQP 133
Db 416 LKCL-----KSLGLT-----GELFEIRLTSLTNADVNTIVKQLSDP 454

QY 134 VAVLEGASDPTSISLSLDKHSQIIN-----K 161
Db 455 V-----ETTKKUTASQTFELDKHSELINTNVSMLETGTRYKKVLSLFTFAIGKGRASS 510

QY 162 FVNSVINTLKTSVSSLOKEICPLRIFHSIDVNVQVVDNPNQHKTOLOTLI 215
Db 511 FLTSFTTLESITFELLRVTPSPAPTALKLISLNIAKYINSIEKEVNIPTLV 564

RESULT 6
F90067
hypothetical protein SA2399 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F90067
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogino, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K. Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F90067
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-296 <KUR>
A:Cross-references: GB:BA000018; PID:g13702563; PIDN:BA843704.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2399

Query Match 9.8%; Score 103.5; DB 2; Length 296;
Best Local Similarity 25.5%; Pred. No. 0.68;
Matches 53; Conservative 30; Mismatches 68; Indels 57; Gaps 10;

QY 8 EGLEVDNTL-KGILEKLVGVLOKSSANOLAKQAEKLLNNVSKLLPTN-TDI 65
Db 83 EGYTADYLDKGVVFFLKVDKGAEQGVQLMKP-----IDNLSLDLRANRHI 134

QY 66 FGLKISNLSILDVKAEPIDGKGLNLSFPVTANVTAG--PIIGQIIN----- 111
Db 135 FGTGM-RSNILENSEQIKD--VVEQGEVAKIAGLVPIIEPVNINAKDKAEIKV 191

QY 112 ----LKASLDLLTA-----VTETDPTQHPVAVLEGASDPTSISL----- 149
Db 192 LKAEKLGKGLDLSNADQLVNLKLTIPTEPNLYK-----ELAEHPNVVRVYVLSGGYSREK 245

QY 150 --SLLDKHSQIINKFVNSVINTLKSFVS 175
Db 246 ANELLKNDIELIASFSRALLASDLRADQS 273

RESULT 7
T50073
myosin-like coiled-coil protein sp8 [imported] - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: T50073
R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Harris, D. submitted to the EMBL Data Library, December 1999
A:Reference number: Z25034
A:Accession: T50073
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1727 <MCD>
A:Cross-references: EMBL:AL133357; PIDN:CAB62414.1; GSPDB:GN000066; SPDB:SPAC1486.04;
A:Experimental source: strain 972h(-); cosmid ci486
C:Genetics:
A:Gene: SPDB:SPAC1486.04C
A:Map position: 1

Query Match 9.4%; Score 99.5; DB 2; Length 1727;
Best Local Similarity 26.1%; Pred. No. 14;
Matches 49; Conservative 36; Mismatches 74; Indels 29; Gaps 8;

QY 23 KLVKLVGVLOKSSANOLAKQ-----KAQAEKLLNNVSKLLPTNTDIFGLKISNLSLI 77
Db 404 KLODLRLTKRFSFCQVKQRIIPVVKQORSEIVRNNTYNNFLSES-----LETSSNNLTK 458

QY 78 VKAEPIDGKGLNLSF-PVTANVTAGPIIGQIINKKASLDLLTAVTIFDPTQHPVAV 136
Db 459 VQAEILLTKMRQEAQVQLTASRTOCSLDLSREVICLMAELDLHNETKSRNVPATVQ--VA 516

QY 137 LGEACSDPTSISLSLDKHSQIINKFVNSVINTLKSTVSSLOKEICPLRIFHSI--D 194
Db 517 LDYANPSTASTELVKN--ELAN-----FSSIKRAVSKTLE-----LREKVALECD 562

QY 195 VNVVQGV 202
Db 563 VEIQKQTV 570

RESULT 8
WMTM8T
180K protein - tomato mosaic virus (strain L)
N:Contains: 130K protein
C:Species: tomato mosaic virus
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 30-Sep-1993
C:Accession: A04195
R:Ohno, T.; Aoyagi, M.; Yamanashi, Y.; Saito, H.; Ikawa, S.; Meshi, T.; Okada, Y. J. Biochem. 96, 1915-1923, 1984
A:Title: Nucleotide sequence of the tobacco mosaic virus (tomato strain) genome and
A:Reference number: A91984; MUID:85157522
A:Accession: A04195
A:Molecule type: genomic RNA
A:Residues: 1-1615 <OHN>
A>Note: readthrough of the terminator UAG between codons for Gln-1116 and Gln-1117
C:Superfamily: cucumber mosaic virus RNA 1 protein

Query Match 9.3%; Score 98.5; DB 1; Length 1615;
Best Local Similarity 20.4%; Pred. No. 16;
Matches 40; Conservative 44; Mismatches 79; Indels 33; Gaps 7;

QY 23 KLVKLVGVLOKSSANOLAKQAEKLLNNVSKLLPTNTDIFGLKISNLSLIIDVKAEP 82
Db 536 KMSVDMPLVD-----IRKKMEETEEMYNALSELVSKNSDKFDDVDFVSCMQSLEVDP 588

C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence revision 18-Aug-1995 #text change 16-Jul-1999

C:Species: Saccharomyces cerevisiae
C:Date: 19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 24-Sep-1999
C:Accession: S57596; S72446
R:Skellton, J.; Church, C.M.
submitted to the EMBL Data Library, June 1995
A:Reference number: S57587
A:Accession: S57596
A:Molecule type: DNA
A:Residues: 1-1729 <SKE>
A:Cross-references: EMBL:249939; NID:g887599; PIDN:CAA90200.1; PID:g887610; MIPS:YMR229c
A:Experimental source: strain AB972
R:Venema, J.; Tollervoy, D.
EMBO J. 15, 5701-5714, 1996
A:Title: RRP5 is required for formation of both 18S and 5.8S rRNA in yeast.
A:Reference number: S72446; MUID:97051828
A:Accession: S72446
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1729 <VEN>
A:Cross-references: EMBL:249939; NID:g887599; PIDN:CAA90200.1; PID:g887610
C:Genetics:
A:Gene: SGD:FWI1; RRP5
A:Cross-references: SGD:S0004842; MIPS:YMR229c
A:Map position: 13R
C:Function:
A:Description: required for pre-rRNA processing
C:Superfamily: ribosomal RNA processing protein RRP5
C:Keywords: nucleus; RNA binding

Query Match 8.9%; Score 94.5; DB 2; Length 1729;
Best Local Similarity 20.5%; Pred. No. 36;
Matches 54; Conservative 44; Mismatches 87; Indels 79; Gaps 13;
QY 11 ETVDNTLKG-----ILEKIK-----VDLGVLOKSSANQLAKQ--KAQE 46
Db 587 DTIENIVPGKTIITVHVLEKTDSDVIEIPDVGLKGVITVGHLSDSRIEQNRAQLKKUR| 746
QY 47 AEKLLNNVTSKILPTNTDIFGLKISNLSILDVKAEPJ-----DDGKGLNLSFPPVTAN---- 98
Db 747 GTELTGLVLDK--DTRTRVFNNSLSKSLKDAKKTPLTYDDVKDINKDVEMHAYIKS| 804
QY 99 -----VTVAGPIIGOTIINLKA-----SLDL-----LTAVTITFD----- 127
Db 805 SDKGLFVAFNGKFIGLVLSYAVDSRDIDISKAFYINOSVTYVILKTDKKNKQKFLLSLKA 864
QY 128 PQTHQVAVIGKASDPTISLSLHKHSQIINKFVNSVINTLKSTVASLQKREICPLIR 187
Db 865 PKVKEKKKVESLHPVDSSIKSDDL-----IGSIV---KAKIKSVKKKNQNLNVILA 915
QY 188 IFIHS-LDV-----NVIQVVVDNPO 206
Db 916 ANLHGRVOIAEVFDYEEITDKKQ 939

Search completed: August 6, 2002, 17:05:14
Job time: 916 sec

Wed Aug 7 05:47:01 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model)

Run on: August 6, 2002, 17:07:32 ; Search time 32.88 Seconds
(without alignments)
253.184 Million cell updates/sec

Title: US-10-020-139-2_COPY_35_249

Perfect score: 1061
Sequence: 1 KLEPVLHGLETVDTNLKGI.....NVIOQVVDNPOHKTQLOTLI 215

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swissprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	294.5	27.8	235	1	PSP_MOUSE
2	185.5	17.5	278	1	PLUN_MOUSE
3	178	16.8	256	1	PLUN_HUMAN
4	106	10.0	295	1	ALF_STACA
5	106	10.0	1769	1	YK9_YEAST
6	99.5	9.4	1727	1	ALM1_SCHPO
7	98.5	9.3	1616	1	RRPO_TOMK2
8	98.5	9.3	1616	1	RRPO_TOML
9	98.5	9.3	1616	1	RRPO_TOMK1
10	95	9.0	679	1	DNLI_HAEIN
11	95	9.0	1531	1	YQ38_CAEEL
12	94.5	8.9	1729	1	RRP5_YEAST
13	93.5	8.8	1038	1	YK03_YEAST
14	92.5	8.7	958	1	YK07_YEAST
15	92.5	8.7	1616	1	RRPO_TOMK1
16	92	8.7	1409	1	HAP1_HAEIN
17	91	8.6	2710	1	TOKA_CLODI
18	90	8.5	529	1	VGUF_SVS
19	89	8.4	368	1	ISPG_LISMO
20	89	8.4	481	1	LBP_RAT
21	89	8.4	1057	1	EG5_HUMAN
22	88	8.3	182	1	PMFF_PROMI
23	87.5	8.2	602	1	ISPG_CHLTR
24	87	8.2	2230	1	GOGA_METTL
25	86.5	8.2	338	1	RLAO_METTL
26	86.5	8.2	668	1	PBPC_BACSU
27	86	8.1	552	1	Y55L_SYNY3
28	85.5	8.1	306	1	DDL_BUCAP
29	85.5	8.1	384	1	DGTP_RICCN
30	84.5	8.0	355	1	CAPH_STAAU
31	84	7.9	1048	1	ITAV_HUMAN
32	83.5	7.9	450	1	TBA3_DROME
33	83.5	7.9	462	1	NIFK_METMP

34 83.5 7.9 615 1 YBM2_SCHPO
35 83.5 7.9 868 1 N180_YEAST
36 83.5 7.9 1301 1 SAC3_YEAST
37 83 7.8 496 1 DIAL_CLAHE
38 83 7.8 699 1 EFG_AGRTO
39 82.5 7.8 466 1 FLID_SALTY
40 82.5 7.8 757 1 DNMI_YEAST
41 82.5 7.8 2144 1 BP28_HUMAN
42 82 7.7 366 1 YA32_METJA
43 82 7.7 444 1 YHDP_BACSU
44 81.5 7.7 444 1 TBA_ONCKE
45 81.5 7.7 450 1 TBA3_MOUSE

ALIGNMENTS

RESULT 1

PSP_MOUSE
ID PSP_MOUSE STANDARD; PRT; 235 AA.
AC P07743;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Parotid secretory protein precursor (PSP).
GN PSP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Parotid gland;
RX MEDLINE=85215456; PubMed=2582349;
RA Madsen H.O., Hjorth J.P.;
RT "Molecular cloning of mouse PSP mRNA."
RL Nucleic Acids Res. 13:1-13(1985).
RN [2]
RP SEQUENCE OF 1-87 FROM N.A.
RC STRAIN=C3H; TISSUE=Splice;
RX MEDLINE=87004556; PubMed=2428613;
RA Poulsen K., Jakobsen B.K., Mikkelsen B.M., Harnmark K.,
RA Nielsen J.T., Hjorth J.P.;
RT "Coordination of murine parotid secretory protein and salivary
amylase expression."
RL EMBO J. 5:1891-1896(1986).
CC -!- FUNCTION: PSP IS THE MOST ABUNDANT PROTEIN IN THE PAROTID GLAND.
CC ITS FUNCTION IS NOT KNOWN; HOWEVER, ITS PRODUCTION IS COORDINATED
CC WITH THAT OF SALIVARY AMYLASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X01697; CAA25846.1;
DR EMBL: M26807; AAA40009.1;
DR EMBL: M26806; AAA40009.1; JOINED.
DR PIR: A23031; SQMS.
DR MGD: MGI:97787; PSP.
DR Parotid gland; Signal.
FT CHAIN 1 20
FT SIGNAL 21 235
FT POTENTIAL.
FT PAROTID SECRETORY PROTEIN.
SQ SEQUENCE 235 AA; 24753 MW; 23311BAE1E5E2EF3 CRC64;

Query Match 27.8%; Score 294.5; DB 1; Length 235;
Best Local Similarity 32.7%; Pred. No. 6.5e-17;
Matches 69; Conservative 55; Mismatches 74; Indels 13; Gaps 4;

Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
Jain J., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
Beasley O.P., Bird C.P., Blake S.E., Bridgeman A.M., Brown A.J.,
Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,
Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
Coulson A., Coville G.J., Deadman R., Dhani P., Dunn M.,
Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
Hammond S., Harley J.R., Heath P.D., Ho S., Holden J.L., Howden P.J.,
Huckle E., Hunt A.L., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
Lehvaslahti M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D.,
Marsh V.L., Martin S.J., McConachie L.J., McLay K., McMurray A.A.,
Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
Skuce C.D., Smith M.L., Soderlund C., Stewart C.A., Sulston J.E.,
Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
Tracey A.C., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
Rogers J.,
"The DNA sequence and comparative analysis of human chromosome 20.";
Nature 414:865-871(2001).
[5]
SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RP Strausberg R.;
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
[6]
SEQUENCE OF 95-152: 157-167 AND 214-232.
RX MEDLINE=21317946; PubMed=11425234;
RT Lindahl M., Stahlboom B., Tagesson C.;
RT "Identification of a new potential airway irritation marker, palate
lung nasal epithelial clone protein, in human nasal lavage fluid with
two-dimensional electrophoresis and matrix-assisted laser
desorption/ionization-time of flight.";
RT Electrophoresis 22:1795-1800(2001).
CC -1- FUNCTION: May be involved in the airway inflammatory response
CC after exposure to irritants. May be associated with tumor
CC progression.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Upper airways and nasopharyngeal regions,
CC including trachea and nasal epithelium. Expressed in lung cancers
CC and some other types of cancer.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AF214562; AAG13653.1; -
DR EMBL: AF172993; AAF70860.1; -
DR EMBL: AB024937; BAA93633.1; -
DR EMBL: AF158745; AAF82622.1; -
DR EMBL: AL121901; AAC03549.1; -
DR EMBL: BC012549; AAH12549.1; -
KW Signal.
FT CHAIN 1 19 POTENTIAL.
FT FT CHAIN 20 256 PROTEIN PLUNC.
FT CONFLICT 220 220 Q -> K (IN REF. 1: AAF70860).
SQ SEQUENCE 256 AA: 26712 MW: EDF152FBC33515BC CRC64;

Query Match 16.8% Score 178; DB 1; Length 256;
Best Local Similarity 24.5% Pred. No. 1.8e-07;
Matches 46; Conservative 44; Mismatches 92; Indels 6; Gaps

```

Query Match      10.0%  Score 106;  DB 1;  Length 1769;
Best Local Similarity 19.2%  Pred. No. 1.4;
Matches 45;  Conservative 44;  Mismatches 73;  Indels 72;  Gaps 8;

QY 23 KUKVDGLGVQKSSAWLAKQKQAEKLLNNVI--SKLLPTNTDI-----FGIKLSNLI 75
|| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

Query Match          9.3%; Score 98.5; DB 1; Length 1616;
Best Local Similarity 20.4%; Pred. No. 4, 9;
Matches 40; Conservative 44; Mismatches 79; Indels 33; Gaps 7;

23 KLVYDGVGVQKSSAWLAKOAKQAEPKLLNNVISKLLPTNTDIFGLKINSLLIDVKAEP 82
    :|||:|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
536 KMSVDMPVLVD-----IRKKMEETEEMYNALSELNLSVKNSDKFDVFQSCOSLEVPD 588
    :|||:|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
83 IDUGK-----GNLISF--PVTANVTAGPIIGIILNKASLDLTAVTETIETDPOT 130
    :|||:|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
589 MTAKVIVAVMKNESGLTLTFEPTCAVALA-----LODEKASDGALVVTSDV 639
    :|||:|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
131 HQPVAVLGECASDFTSIS--LSLLDKHSOIINKFNVSINTLKSTVSSLQKEICPLRI 188
    :|||:|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
640 EEP-SIRGSWARGELQLAGLGDVPPESSYTRSEETLESQEHFMAATASSLIHKQMCSI-V 696
    :|||:|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
189 FIHSLDVNIQQVVVDN 204
    :|||:|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
697 YTGPKVQVMKNFIDS 712
    :|||:|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

RESULT      8
RPRO_TOML
AC RPO_TOML STANDARD; PRT; 1616 AA.
AD P03587; O41352;
ID 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DD 16-OCT-2001 (Rel. 40, Last annotation update)
DE Methy-directed RNA polymerase (EC 2.7.7.48) (183 kDa protein) [Contains:
DE   Methytransferase/RNA helicase (MT/HEL) (126 kDa protein)].
DE Tomato mosaic virus (strain L) (TOMV) (TMV strain tomato).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=12252;
[1]
SEQUENCE FROM N.A.
MEDLINE=85157522; PubMed=6549393;
RA Ohno T., Aoyagi M., Yamashita Y., Salto H., Ikawa S., Meshi T.,
RA Okada Y.;
RA "Nucleotide sequence of the tobacco mosaic virus (tomato strain)
RA genome and comparison with the common strain genome.";
RRL J. Biochem. 96:1915-1923(1984).
RRT -! FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
RRT RNA REPLICATION.
CCC -! FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CCC CAPPING AND AN RNA HELICASE.
CCC -! MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CCC BETWEEN COLUMNS FOR GLN-1116 AND GLN-1118.
CCC -----
CCC This SWISS-PROT entry is copyright. It is produced through a collaboration
CCC between the Swiss Institute of Bioinformatics and the EMBL outstation
CCC at the European Bioinformatics Institute. There are no restrictions on its
CCC use by non-profit institutions as long as its content is in no way
CCC modified and this statement is not removed. Usage by and for commercial
CCC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CCC or send an email to license@isb-sib.ch).
CCC -----
DR EMBL; X02144; CA26085.1; -
DR EMBL; X02144; CA26082.1; -
DR PIR; A04195; WMTM81.
DR InterPro; IPR001788; RNA_dep_RNApol2.
DR InterPro; IPR002588; V_methyltransf.
DR InterPro; IPR000606; Viral_helicasel.
DR Pfam; PF00978; RNA_dep_RNApol2; 1.
DR Pfam; PF01443; Viral_helicasel; 1.
DR Pfam; PF01660; Vmethyltransf; 1.
DR Transferrase; RNA-directed RNA polymerase; Helicase; ATP-binding.
KW CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE
FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 833 840 ATP (POTENTIAL).
FT SEQUENCE 1616 AA; 103564 MW; A8EC8929B5CF7CAF CRC64;
SQ

```

Query Match 9.38; Score 98.5; DB 1; Length 1616;

DB QY DB
536 KMSVMPVLQ-----IRKKMEETEMYNALSELNLSKDKFDVDVFQMCCSLSEVPD 586

DB QY DB
83 IDDGK-----GLNISF--PVTAVTVAGPTIGGIINLKASLULLTAVTITDPOT 130

DB DB
589 MTAARVVIVAVNSGSLUTLFEGOPTAANVALA-----LDSEKASDGAUVVTSKOV 639

DB QY DB
131 HQPVAVLGECASDPTSTS--LSLUKHSHQIINKFVNSVINTLKSTVSLLQKEICPLIRI 188

DB DB
640 EEP-SIKSGMARGELQLAGLSDGVPESSVTRSEETESLEQFHMTASSLHKMCST--V 696

DB QY DB
189 FHSLDNVNIQQVDN 204

DB DB
697 YTGPLKVQQMKNFIDS 712

RESULT 10

ID	DNLJ HAEIN	STANDARD;	PRT:	679 AA.
AC	P43813;			
AT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DD	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	DNA ligase (BC 6.5.1.2) (Polydeoxyribonucleotide synthase [NAD+])			
GN	LIGN OR Lig OR Ht1100.			
OS	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;			
OC	Haemophilus			
ON	NCB1_TaxID=727;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-RD / KW20 / ATCC 51907;			
RX	MEDLINE=95350630; PubMed=7542800;			
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,			
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,			
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,			
RA	Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.W.,			
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,			
RA	Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,			
RA	Fine L.D., Frithman J.L., Fuhrmann J.L., Geoghagen N.S.M.,			
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,			
RA	Venter J.C.;			
RT	"Whole-genome random sequencing and assembly of Haemophilus			
RL	influenzae Rd."			
RL	Science 269:496-512(1995).			
CC	-!- FUNCTION: THIS PROTEIN CATALYZES THE FORMATION OF PHOSPHODIESTER			
CC	LINKAGES BETWEEN 5'-PHOSPHORYL AND 3'-HYDROXYL GROUPS IN DOUBLE-			
CC	STRANDED DNA USING NAD AS A COENZYME AND AS THE ENERGY SOURCE FOR			
CC	THE REACTION. IT IS ESSENTIAL FOR DNA REPLICATION AND REPAIR OF			
CC	DAMAGED DNA (BY SIMILARITY).			
CC	-!- CATALYTIC ACTIVITY: NAD(+) + [deoxyribonucleotide](N) +			
CC	[deoxyribonucleotide](N)= AMP + nicotinamide nucleotide +			
CC	[deoxyribonucleotide](N)+H ₂ O			
CC	-!- SIMILARITY: BELONGS TO THE NAD-DEPENDENT DNA LIGASE FAMILY.			
CC	-!- SIMILARITY: CONTAINS 1 BRCT DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-			
CC	tion between the Swiss Institute of Bioinformatics and the EMBL Outstation			
CC	at the European Bioinformatics Institute. There are no restrictions on the			
CC	use by non-profit institutions as long as its content is in the public			
CC	domain and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/about/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: U32789; AAC22753.1; ..			
DR	HSP: 087703; IB04.			
DR	TIGR: HT1100; ..			
DR	InterPro: IPR001357; BRCT			
DR	InterPro: IPR001679; DNA_ligase_N.			
DR	InterPro: IPR004150; DNA_ligase_OB.			
DR	InterPro: IPR004149; DNA_ligase_2BD.			
DR	InterPro: IPR000445; HHH.			
DR	InterPro: IPR003583; HHH.			

DR PFam: PF00533; BKCT: 1.
DR PFam: PF01653; DNA_ligase_N; 1.
DR PFam: PF01120; DNA_ligase_OB; 1.
DR PFam: PF03119; DNA_ligase_ZBD; 1.
DR PFam: PF00633; HHH; 1.
DR ProDom: PD003944; DNA_ligase_N; 1.
DR SMART: SM00292; BKCT: 1.
DR SMART: SM00278; HHH; 2.
DR SMART: SM00532; LIGANC; 1.
DR PROSITE: PS0172; BKCT: 1.
DR PROSITE: PS01055; DNA_LIGASE_N1; 1.
DR PROSITE: PS01056; DNA_LIGASE_N2; 1.
KW Ligase; DNA repair; DNA replication; NAD; Complete proteome.
FT DOMAIN 601 679 BKCT.
FT BINDING 125 125 AMP (BY SIMILARITY).
FT SEQUENCE 679 AA; 75188 MW; 4C0C23E25D70FF52 CRC64;
Query Match 9.0%; Score 95; DB 1; Length 679;
Best Local Similarity 24.5%; Pred. No. 3.1;
Matches 52; Conservative 40; Mismatches 68; Indels 52; Gaps 11;
QY 13 VONTLKIGLE-KLKVDGLGVLOKSSAWQLA-KOKAQEAELNNV-----ISKL 58
DB 294 IDGTVLKINDIALQNELGFISKAPWATAYKFPAGEELTLNDVEFGVGTGATPPAKL 353
QY 59 LPTNTDIFGLKISNLSLDVKAEPIDGKGLNLSFPVTANTVAGPIIGQIINLKASLDL 118
DB 354 EP--VFVAGTVSNATLHNG-----DETERLNIAIGDVFWRAGDVIPQII----- 398
QY 119 LTAVTIETDPTQHPVAVLGECASTPTISLSLDKHSQIKFNVSINTLKTSTVSL 178
DB 399 --GVLHERPDNAKPIIF-----PTNCPVC---DSQIRIEGEAVAKTUGLFCQAQ 445
QY 179 QKEICPLIRIFI--HSLDYN-----VIOQVVD 203
DB 446 RKE---ALKHEVSRKAMDIDGVGKXLEQLVD 474
RESULT 11
ID YQ38-CABEL STANDARD; PRT: 1531 AA.
AC Q09459; Q09513;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical 174.9 kDa protein C0965.8 in chromosome II.
GN C0965.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Palmer S., Sulston J.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO C.ELEGANS 24666.1 AND 24666.4
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 246791; CAA86762.2;
CC EMBL; 246792; CAA86762.2; JOINED.

DR EMBL; 246792; CAA86767.2;
DR EMBL; 246791; CAA86767.2; JOINED.
DR WormPep: C0965.8; C252774.
KW Hypothetical protein. POLY-SER.
FT DOMAIN 384 394
FT DOMAIN 395 398 POLY-ASN.
FT DOMAIN 1150 1154 POLY-ASP.
FT SEQUENCE 1531 AA; 174862 MW; A453990AF0DEBC2F CRC64;
Query Match 9.0%; Score 95; DB 1; Length 1531;
Best Local Similarity 24.9%; Pred. No. 8.7;
Matches 65; Conservative 33; Mismatches 73; Indels 90; Gaps 13;
QY 1 KLEPVLHEGLETVDN-----TLKGILEKLA-VDLGVLOKSSAWQLAKQ 42
DB 648 KLYSELYEELEKVRNMLLIQYDINOKMKKEIKLLKDELDRKLTSAEILSKSREEVEERQ 707
QY 43 K-----AOEAELNNVSKLLPT---NTD-----IFGLKISNLSL----- 75
DB 708 KKIEMLEEQIRTIAYSGQQPQVKKLANQIN--IPTPRVNTDLSVKLINVKPSPSLTSKFF 765
QY 76 -----LDVKAEPIDGKGLNLSFPVTANTVAGPIIGQIINLKASLDLTAVTIETDPQ 129
DB 766 SLEFFDFQLETPIMDAKQHNMDFTTVYDVLVSNLLIHYL-----Q 806
QY 130 THOPVAVLGECAST-----PTISLSLDKHSQIKFNVSINTLKTSTVSLLOKEIC 183
DB 807 TNGIVITMYPASDCYKLLAAATISLIPLFEDS-VLRKPCSEIM--LKSVDGV---EMC 860
QY 184 PLIRIFIHSLDYNVVIQVVDN 204
DB 861 TL-----RYEIEVSQIPISDS 875
RESULT 12
ID RRP5-YEAST STANDARD; PRT: 1729 AA.
AC Q05022;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE rRNA biogenesis protein RRP5.
GN RRP5 OR FM11 OR YMR229C OR YMR959.11C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB372;
RA Skellton J., Churcher C.M., Bartell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP CHARACTERIZATION.
RX MEDLINE-97051828; PubMed-8896463;
RA Venema J., Tollervay D.;
RT *RRP5 is required for formation of both 18S and 5.8S rRNA in yeast.*
RL EMBL J. 15:5701-5714(1996).
CC -1- FUNCTION: INVOLVED IN THE BIOGENESIS OF RNA. REQUIRED FOR THE
CC FORMATION OF 18S AND 5.8S RNA.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR.
CC -1- SIMILARITY: CONTAINS 4 HAT REPEATS.
CC -1- SIMILARITY: CONTAINS 11 SI MOTIF DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 246791; CAA86762.2;
CC EMBL; 246792; CAA86762.2; JOINED.

protein and a homologue to the yeast mitochondrial regulator ABF2.:"

RT EMBL: 249939; CAA90200.1; --
 RL HSP; P05055; 1SKO.
 RN SGP; S0004842; RRP5.
 RP InterPro: IPR003107; RNT.
 RA Pfam: PF00575; S1; 7.
 DR SMART: SMO0386; HAT; 2.
 DR SMART: SMO0316; S1; 12.
 DR PROSITE: PS0126; S1; 11.
 KW Nuclear protein; rRNA processing; Repeat.
 FT DOMAIN 119 200
 FT DOMAIN 338 410
 FT DOMAIN 510 580
 FT DOMAIN 507 676
 FT DOMAIN 690 769
 FT DOMAIN 794 863
 FT DOMAIN 895 971
 FT DOMAIN 1003 1083
 FT DOMAIN 1088 1159
 FT DOMAIN 1177 1245
 FT DOMAIN 1265 1336
 FT REPEAT 1455 1487
 FT REPEAT 1561 1594
 FT REPEAT 1632 1664
 FT REPEAT 1666 1701
 FT REPEAT 1729 1799
 SQ SEQUENCE 1729 AA; 193133 MW; 39BF465587B3B0A CRC64;

Query Match 8.9%; Score 94.5; DB 1; Length 1729;
 Best Local Similarity 20.5%; Pred. No. 11;
 Matches 54; Conservative 44; Mismatches 87; Indels 79; Gaps 13;

QY 11 ETVDNTLKG-----ILEKLLK-----VDLGVLOKSSAWQLAKQ--KAQE 46
 DB 687 DTIIIVPGRTITIVHVEIKTRDSVTEIPDVLGRGVIVGHLSDSRIEQNRAQLKLRI 746
 QY 47 AEKLLNNVSKLPTNTDIFGLKISNLSLDVKAEPF-----DDGKGLNLSFPTAN----- 98
 DB 747 GTELTGLVIDK--DTRTRVFNMSSKSLKDAKKTPLTYDDVDKLDKDVPMHAYIKSI 804
 QY 99 -----TVVAGPIIGOTINLKA-----SLDL-----LTAVTIETD----- 127
 DB 805 SDKGLFVAFNGKFIGLVPSYAVDSRDDIISKAFYINQSVTVYLLRTDDKDKFKLLSLKA 864
 QY 128 PQHQPVAVLGCASDPTISLSLLDKHSQINLKFNKNSVINTLKSTVSSLLKKEICPLIR 187
 DB 865 PKVKEKKKVESNIEDPVUSSIKSMDULS-----IGSIV---KAKIKSVKKKNQINVLIA 915
 QY 188 IFHS-LDV-----NVIQVVVDNQ 206
 DB 916 ANLHGRVDIAEVFDYEEITDKKQ 939

RESULT 13
 YK03_YEAST
 ID YK03_YEAST STANDARD; PRT; 1038 AA.
 AC P36097;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE Hypothetical 118.9 kDa protein in PTM1-IXR1 intergenic region.
 GN YK033W OR YK0246.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=94262309; PubMed=8203146;
 RA Punnelle B., Skala J., van Dyck L., Goffeau A.;
 RT "Analysis of an 11.7 kb DNA fragment of chromosome XI reveals a new
 tRNA gene and four new open reading frames including a leucine zipper

protein and a homologue to the yeast mitochondrial regulator ABF2.:"

RT EMBL: 249939; CAA90200.1; --
 RL HSP; P05055; 1SKO.
 RN SGP; S0004842; RRP5.
 RP InterPro: IPR003107; RNT.
 RA Pfam: PF00575; S1; 7.
 DR SMART: SMO0386; HAT; 2.
 DR SMART: SMO0316; S1; 12.
 DR PROSITE: PS0126; S1; 11.
 KW Nuclear protein; rRNA processing; Repeat.
 FT DOMAIN 119 200
 FT DOMAIN 338 410
 FT DOMAIN 510 580
 FT DOMAIN 507 676
 FT DOMAIN 690 769
 FT DOMAIN 794 863
 FT DOMAIN 895 971
 FT DOMAIN 1003 1083
 FT DOMAIN 1088 1159
 FT DOMAIN 1177 1245
 FT DOMAIN 1265 1336
 FT REPEAT 1455 1487
 FT REPEAT 1561 1594
 FT REPEAT 1632 1664
 FT REPEAT 1666 1701
 FT REPEAT 1729 1799
 SQ SEQUENCE 1729 AA; 193133 MW; 39BF465587B3B0A CRC64;

Query Match 8.8%; Score 93.5; DB 1; Length 1038;
 Best Local Similarity 21.4%; Pred. No. 7;
 Matches 58; Conservative 46; Mismatches 80; Indels 87; Gaps 12;

QY 5 VLHEGLTV-----DNTLKGILEKLVGLGVLOKSSAWQLAKQA-QEAEK 49
 DB 141 VLHQFFKSVRSQRYHKEFFSNKPNILPALGHSVTILLKLSQSPQNDELQFKALASLEV 200
 QY 50 LLANNVSK-----LLPTNTDIFG-----LKISNLSI-----LD 77
 DB 201 LFOQISDGEMLSFILPGNVSVFAKILTKPGRIHYKVCVRLVLEAKLLVLVYDDFSLD 260
 QY 78 VKAEPIDGGKGLN---LSFPVTANTVAGPIIQIINLKASLDLLTAVTIETDPTQHPV 134
 DB 261 IKVKNKLTDIRLSDTKLKHEINQSFNNGPI-----VLLRTDGTGTHROT 304
 QY 135 AVLCECASDPTISLSL-----SLDKHSQIINK-----FV-----NSVINTLKSTVSSL 177
 DB 405 SWL-TATSGUINIALEAFIPKLLKRNESIDEALATFVILLTRCENSLNCKEVLVSTL 363
 QY 178 IQKEICPLIRIFTHSLDVNVIQVVVDNQPHK 208
 DB 364 VHLEKDEPMKLPShLVK---LKEVNVNEDLHK 391

RESULT 14
 YK07_YEAST
 ID YK07_YEAST STANDARD; PRT; 958 AA.
 AC P53076;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 108.2 kDa protein in SAP4-OST5 intergenic region.
 GN YK0227W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Fatmann B., Kramer B., Kramer W.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: 272749; CAA96943.1; -
 SGB: 50003196; VID30.
 InterPro: IPR003877; SPRY.
 InterPro: IPR003878; SPRY_domain.
 Pfam: PF00622; SPRY; 1.
 SMART: SM00449; SPRY; 1.
 KW Hypothetical protein.
 FT DOMAIN 165 185 POLY-ASP.
 FT DOMAIN 450 453 POLY-SER.
 SEQUENCE 958 AA; 108178 MW; 335ADDJ52949F8C8 CRC64;

Query Match 8.7%; Score 92.5; DB 1; Length 958;
 Best Local Similarity 22.8%; Pred. No. 7.6;
 Matches 51; Conservative 39; Mismatches 71; Indels 63; Gaps 11;

QY 5 VLHEGLETVDTNLKGLLEKIAVDL-----GVLOKSSAWOLAKQK-----AQEAE 48
 DB 722 LIHEGLVDV---AKGLKDLQKDAVNVGCHSEKDVIRHNER-QIMKEEMVKIROELR 777
 QY 49 KLIN-----NVISKLLPTNTD-IFGLKISNSLIDVKAEPIDGG-----KG 88
 DB 778 YLIMGQISKICINVIDNEIPDLKNNLPELKLANYLVMIKSSKDDDEIENLILKG 837
 QY 89 LNL5-----FPVTVANTVAGPIIGIINLKASLDLLTAVTIEDTPQTHOPVAVLGECASD- 143
 DB 938 QELNSFEIYDTKIPQSLDRFRSGSLNVSALL-----AYSNPLVEAPKEISGYLSDEY 890
 QY 144 -----PTSLSLLDKHSQIINKFNVSINTLKSTVSSLLQ 179
 DB 891 LQERLFOVSNNTILTLFLHKDSECA---LENVISNTRAMLSTLLE 931

RESULT 15
 RRPO_TOMK1 STANDARD; PRT: 1616 AA.
 AC Q901T8; Q901T7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE RNA-directed RNA polymerase (EC 2.7.7.48) (183 kDa protein) [Contains:
 DE Methyltransferase/RNA helicase (MI/HEL) (126 kDa protein)]
 CS Tomato mosaic virus (strain Kazakh K1) (ToMV) (TMV strain K1).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
 OX NCBI_TaxID=138311;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196905; PubMed=10732356;
 RA Belenovich E.V., Novikov V.K., Zavrjev S.K.;
 RT "Biological properties and genome structure of the Kazakh isolate K1
 of Tobacco Mosaic virus."
 RL Mol. Biol. (Mosk) 34:172-176(2000).
 CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
 RNA REPLICATION.
 CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
 CAPPING AND AN RNA HELICASE.
 CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
 BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AJ243571; CAB62911.1; -
 DR EMBL: AJ243571; CAB62912.1; -
 DR InterPro: IPR001788; RNA_dep_RNAPol2.
 DR InterPro: IPR002588; V_methyltransf.
 DR InterPro: IPR000606; Viral_helicase.
 DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
 DR Pfam: PF01443; Viral_helicase; 1.
 DR Pfam: PF01660; Vmethyltransf; 1.
 KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
 FT CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
 FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
 FT NP_BIND 833 840 ATP (POTENTIAL).
 SEQUENCE 1616 AA; 183606 MW; 4BF9913CE79IDE33 CRC64;

Query Match 8.7%; Score 92.5; DB 1; Length 1616;
 Best Local Similarity 19.9%; Pred. No. 15;
 Matches 39; Conservative 44; Mismatches 80; Indels 33; Gaps 7;

QY 23 KLVVDLGVLOKSSAWOLAKQAKAEKLLNNVSKLLPTNTDIFGLKISNSLIDVKAEP 82
 DB 536 KMSVDMPVLD-----IRKMEETEEMYNALSELVSLKTSKDFDQVDFSQMCQSLVDP 588
 QY 83 IDDGK-----GLNLSF--PVTANTVAGPIIGIINLKASLDLLTAVTIEDTPQ 130
 DB 589 MTAAKVIVAVMNSGSLTLTPEQTEANVALA-----LDSEKASDGLAVTSDV 639
 QY 131 HQPVAVLGECASDPTIS--LSLLDKHSQIINKFNVSINTLKSTVSSLLQKEICPLIRI 188
 DB 640 EEP-SIRGSMARGELOLQAGLSDGVPPESSYTRSEEIESLEQFHMATASSLIHKQMCST--V 696
 QY 189 FTHSIDVNVIIQOVVN 204
 DB 697 YTGPLKVQQMKNFIDS 712

Search completed: August 6, 2002, 17:07:34
 Job time: 876 sec

us-10-020-139-2_copy_35_249.rsp

Wed Aug 7 05:47:01 2002

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 17:09:28 ; Search time 111.35 Seconds
(without alignments)
334.027 Million cell updates/sec

Title: US-10-020-139-2_COPY_35_249
Perfect score: 1061
Sequence: 1 KLEPVLHEGLTVDTNLTGK1.....NVIQQVNDPNQKHTQLOTLLI 215

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPTREMBL_19:

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_plant:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1061	100.0	249	4 Q9BQ00	Q9BQ00 homo sapien
2	1053	99.2	249	4 Q9BDR5	Q9BDR5 homo sapien
3	333.5	31.4	243	6 P79124	P79124 bos taurus
4	324	30.5	240	6 P79125	P79125 bos taurus
5	294.5	27.8	235	11 Q9D734	Q9D734 mus musculus
6	263.5	24.8	235	11 Q63471	Q63471 rattus norv
7	154.5	14.6	206	11 Q63550	Q63550 rattus norv
8	148.5	14.0	235	11 Q9D6P2	Q9D6P2 mus musculus
9	148.5	14.0	270	11 Q9D794	Q9D794 mus musculus
10	148.5	14.0	270	11 Q9CQX3	Q9CQX3 mus musculus
11	145.5	13.7	270	11 Q9D6P0	Q9D6P0 mus musculus
12	145	13.7	484	4 Q96HK6	Q96HK6 homo sapien
13	129	12.2	474	11 Q61114	Q61114 mus musculus
14	126	11.9	232	11 Q9D9J8	Q9D9J8 mus musculus
15	115	10.8	199	4 Q9BQ08	Q9BQ08 homo sapien
16	103.5	9.8	296	16 Q99R31	Q99R31 staphylococ

```

17 98.5 9.3 1116 12 Q991T0
18 97.5 9.2 211 17 Q59040
19 97.5 9.2 866 12 Q9WAG5
20 97.5 9.2 1116 12 Q9JA03
21 97.5 9.2 1616 12 Q9JA04
22 94.5 8.9 825 16 Q9JTN0
23 94.5 8.9 1116 12 Q911R8
24 94.5 8.9 1616 12 Q911R9
25 93.5 8.8 2938 11 Q61769
26 93 8.8 551 12 Q9DUT4
27 93 8.8 624 2 Q49548
28 93 8.8 1237 16 Q92CG8
29 92.5 8.7 387 5 Q9XTU2
30 92 8.7 2349 2 P94750
31 92 8.7 2383 16 P76347
32 91.5 8.6 441 11 Q9D749
33 91.5 8.6 441 11 Q9ERV5
34 91.5 8.6 473 11 Q05701
35 91 8.6 443 16 Q9K889
36 91 8.6 454 5 Q17511
37 91 8.6 578 16 Q9K847
38 91 8.6 5233 5 Q9NB71
39 90.5 8.5 673 16 Q9CKA9
40 90.5 8.5 951 10 Q9FVX8
41 90.5 8.5 1254 16 Q9KP32
42 90.5 8.5 1878 2 Q59986
43 90.5 8.5 2752 5 Q9BJY0
44 90 8.5 383 5 Q9N2L8
45 90 8.5 529 12 Q9YZA2

```

ALIGNMENTS

```

RESULT 1
Q9BQ00 PRELIMINARY; PRT; 249 AA.
ID Q9BQ00;
AC Q9BQ00;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE BA49G10.1 (SIMILAR TO BOVINE SALIVARY PROTEIN BSP30).
DE BA49G10.1.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tracey A.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases..
DR EMBL; AL121901; CAC03546.1; .....
SQ SEQUENCE 249 AA; 27011 MW; E64E0794A1B4DB7D CRC64;

```

```

Query Match 100.0%; Score 1061; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 9.2e-78;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLEPVLHEGLTVDTNLTGKLVLDGLVQKSSAWQLAKQAEKLLNNVSKLLP 60
DQ 35 KLEPVLHEGLTVDTNLTGKLVLDGLVQKSSAWQLAKQAEKLLNNVSKLLP 94
QY 61 TNDIFGLKINSILVDKAEPTDDGKGLNSFPVTANVTAGPIIGQIINKASLDLTL 120
DQ 95 TNDIFGLKINSILVDKAEPTDDGKGLNSFPVTANVTAGPIIGQIINKASLDLTL 154
QY 121 AVTIETDPQTHQPVAVLGECASDPTSLSLDKHSGIINKFVNSVINTLKSTVSSLLQK 180
DQ 155 AVTIETDPQTHQPVAVLGECASDPTSLSLDKHSGIINKFVNSVINTLKSTVSSLLQK 214
QY 181 EICPLIRIFHSLDVNVIQQVNDPNQKHTQLOTLLI 215
DQ 181 EICPLIRIFHSLDVNVIQQVNDPNQKHTQLOTLLI 215

```

Db 215 EICPLIRIFHSLDNNVIOQVVDNPNQHKTLQTLI 249

RESULT 2

Q96DR5 PRELIMINARY; PRT; 249 AA.
 AC Q96DR5;
 DT 01-DEC-2001 (TREMELrel. 19, Created)
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
 DE PAROTID SECRETORY PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PAROTID;
 RA Venkatesh S.G., Geetha C., Gorr S.-U.;
 RT "A member of the PSP/plunc family of BPI proteins is expressed in the human parotid gland."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF432917; AAL28113.1; -;
 SQ SEQUENCE 249 AA; 27110 MW; FD54B624A1A4CA7C CRC64;

Query Match 99.2%; Score 1053; DB 4; Length 249;
 Best Local Similarity 99.5%; Pred. No. 4.1e-77;
 Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLEPVHLEGLETVDTNLTGKILEKLVGLQKSSAWOLAKQAEKLLNNVSKLLP 60
 DB 35 KLEPVHLEGLETVDTNLTGKILEKLVGLQKSSAWOLAKQAEKLLNNVSKLLP 94
 QY 61 TNDTIFGLKISNLSLDVKAEPIDDCGKLNLSFPVTANTVAGPIIGQINLKASLDLT 120
 DB 95 TNDTIFGLKISNLSLDVKAEPIDDCGKLNLSFPVTANTVAGPIIGQINLKASLDLT 154
 QY 121 AVTIETDPOTHQPVAVLGCASDPTSISLSLLDKHSQIINKFVNSVINTLKSTVSSLLQK 180
 DB 155 AVTIETDPOTHQPVAVLGCASDPTSISLSLLDKHSQIINKFVNSVINTLKSTVSSLLQK 214
 QY 181 EICPLIRIFHSLDNNVIOQVVDNPNQHKTLQTLI 215
 DB 215 EICPLIRIFHSLDNNVIOQVVDNPNQHKTLQTLI 249

RESULT 3

P79124 PRELIMINARY; PRT; 243 AA.
 AC P79124;
 DT 01-MAY-1997 (TREMELrel. 03, Created)
 DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
 DE BSP30.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PAROTID GLAND;
 RA Haigh B.J., Wilkins R.J., Wheeler T.T.;
 RT "The cloning and sequencing of two cDNAs coding for alternate forms of BSP30, a bovine member of the Parotid Secretory protein family."
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U79413; AAB38282.1; -;
 SQ SEQUENCE 243 AA; 26877 MW; OC2D8DD45660E11C CRC64;

Query Match 31.4%; Score 333.5; DB 6; Length 243;
 Best Local Similarity 34.1%; Pred. No. 3.3e-19;

Matches 71; Conservative 54; Mismatches 78; Indels 5; Gaps 2;
 QY 6 LHEGLETVDNLTGKILEKLVGLQKSSAWOLAKQAEKLLNNVSKLLPNTDI 65
 DB 37 LHEGLETVDNLTGKILEKLVGLQKSSAWOLAKQAEKLLNNVSKLLPNTDI 92
 QY 66 FGLKISNLSLDVKAEPIDDCGKLNLSFPVTANTVAGPIIGQINLKASLDLTAVTIE 125
 DB 93 FGLKISNLSLDVKAEPIDDCGKLNLSFPVTANTVAGPIIGQINLKASLDLTAVTIE 151
 QY 126 TDPOTHQPVAVLGCASDPTSISLSLLDKHSQIINKFVNSVINTLKSTVSSLLQKICPL 185
 DB 152 TDPOTHQPVAVLGCASDPTSISLSLLDKHSQIINKFVNSVINTLKSTVSSLLQKICPL 211
 QY 186 IRIHSLDNNVIOQVVDNPNQHKTLQTL 213
 DB 212 IRIHSLDNNVIOQVVDNPNQHKTLQTL 239

RESULT 4

P79125 PRELIMINARY; PRT; 240 AA.
 AC P79125;
 DT 01-MAY-1997 (TREMELrel. 03, Created)
 DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
 DE BSP30.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PAROTID GLAND;
 RA Haigh B.J., Wilkins R.J., Wheeler T.T.;
 RT "The cloning and sequencing of two cDNAs coding for alternate forms of BSP30, a bovine member of the Parotid Secretory protein family."
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U79414; AAB38283.1; -;
 SQ SEQUENCE 240 AA; 26513 MW; 850611DE9E43E358 CRC64;

Query Match 30.5%; Score 324; DB 6; Length 240;
 Best Local Similarity 33.8%; Pred. No. 1.9e-18;

Matches 71; Conservative 53; Mismatches 78; Indels 8; Gaps 4;
 QY 1 KLEPVHLEGLETVDTNLTGKILEKLVGLQKSSAWOLAKQAEKLLNNVSKLLP 60
 DB 32 KLEPVHLEGLETVDTNLTGKILEKLVGLQKSSAWOLAKQAEKLLNNVSKLLP 86
 QY 61 TNDTIFGLKISNLSLDVKAEPIDDCGKLNLSFPVTANTVAGPIIGQINLKASLDLT 120
 DB 87 TNDTIFGLKISNLSLDVKAEPIDDCGKLNLSFPVTANTVAGPIIGQINLKASLDLT 145
 QY 121 AVTIETDPOTHQPVAVLGCASDPTSISLSLLDKHSQIINKFVNSVINTLKSTVSSLLQK 180
 DB 146 AVTIETDPOTHQPVAVLGCASDPTSISLSLLDKHSQIINKFVNSVINTLKSTVSSLLQK 203
 QY 181 EICPLIRIFHSLDNNVIOQVVDNPNQHKTLQ 210
 DB 204 EICPLIRIFHSLDNNVIOQVVDNPNQHKTLQ 233

RESULT 5

Q9D734 PRELIMINARY; PRT; 235 AA.
 AC Q9D734;
 DT 01-JUN-2001 (TREMELrel. 17, Created)
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
 DE PAROTID SECRETORY PROTEIN.
 GN PSP.

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;
 RX MEDLINE=21055660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Akawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Bozell D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guslinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Willing L.,
 RA Wyushaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK009654; BAB26418.1; -
 DR MGD: MGI:97787; Psp.
 SQ SEQUENCE 235 AA; 24737 MW; A29D0160268DA0CF CRC64;

Query Match 27.8%; Score 294.5; DB 11; Length 235;
 Best Local Similarity 32.7%; Pred. No. 4.3e-16;
 Matches 69; Conservative 55; Mismatches 74; Indels 13; Gaps 4;

OY 3 EPVLRGCTVDNLK-----GILEKLVKVDLCVQKSSAWOLAKQAKAEKLLNNVI 55
 DB 19 ESLICE-LGSVNNKILNPPEAVPNLNDVLLQOATSWPLAKNSILET---LNTAD 74
 OY 56 SKLLPTNTDIFG--LKINSLLIDVKAEPIDGKGLNLSFPVTANTVAGTIIGQIINLK 113
 DB 75 LGNLKFTSLGULLKILNNKLVDFQAKLSNGNCIDLTVPAGEASLVLPFGKTVDIS 134
 OY 114 ASLDLITAVTIFTPQHPQVAVLGECASDPTSLSLDKHSHQIINKFVNSVINTLAKST 173
 DB 135 VSLDLINSLTKTNAQGLPEVTIGKCSNTDKISISLLGRRLPTINSLLDGVSTLTST 194
 OY 174 VSSLLQKEICPLIRIFIHSLDVNVIQGVND 204
 DB 195 LSTVLQNLFLCPLQVLTSLNPSVLOGLSN 225

RESULT 6
 ID Q63471 PRELIMINARY; PRT; 235 AA.
 AC Q63471;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE NEONATAL SUBMANDIBULAR GLAND PROTEIN PRECURSOR.
 GN PSP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SUBMANDIBULAR GLAND;
 RX MEDLINE=92129360; PubMed=1370829;
 RA Mirels L., Ball W.D.;
 RT "Characterization of the rat salivary-gland B1-immunoreactive
 RT proteins";
 RL Biochem. J. 330:437-444(1998).
 DR EMBL: M83210; AAC12783.1; -
 KW Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 206
 FT "Neonatal rat submandibular gland protein SMG-A and parotid secretory

RT protein are alternatively regulated members of a salivary protein
 RL multigene family.";
 RL J. Biol. Chem. 267:2679-2687(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SUBMANDIBULAR GLAND;
 RX MEDLINE=98129760;
 RA Mirels L., Miranda A.J., Ball W.D.;
 RT "Characterization of the rat salivary-gland B1-immunoreactive
 RT proteins";
 RL Biochem. J. 330:437-444(1998).
 DR EMBL: M83209; AAC06334.1; -
 KW Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 235
 FT "NEONATAL SUBMANDIBULAR GLAND PROTEIN.
 SQ SEQUENCE 235 AA; 24529 MW; 0B36EC779025986E CRC64;

Query Match 24.8%; Score 263.5; DB 11; Length 235;
 Best Local Similarity 29.6%; Pred. No. 1.3e-13;
 Matches 59; Conservative 56; Mismatches 75; Indels 9; Gaps 4;

OY 10 LETVDNTLKGILEKLVKVDLCVQKSSAWOLAKQAKAEKLLNNVSKLLPTN--TDIFG 67
 DB 32 LDILNSPEAVQNLNDVGLSLAQATTPSKKDSILET---LNKV--ELGNSNGFTPLNG 86
 OY 68 --LKINSLLIDVKAEPIDGKGLNLSFPVTANTVAGTIIGQIINLKASLDLLTAVTIE 125
 DB 87 LLLRVNKFVLDLQAGLSSNGKIDLKLPLVFEISFSLPGLPTLDVAVSLDLNLSVQ 146
 OY 126 TDPOTHQPVAVLGECASDPTSLSLDKHSHQIINKFVNSVINTLAKSTVSSLLQKEICPL 185
 DB 147 TNAQGLPEVTIGKCSNTDKISISLLGRRLPFVNRILDGVSGLLTGAVSILLQNLCPV 206
 OY 186 TRIFIHSLDVNVIQGVND 204
 DB 207 LQYLLTWSGSAIQGLSN 225

RESULT 7
 ID Q63550 PRELIMINARY; PRT; 206 AA.
 AC Q63550;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE NEONATAL SUBMANDIBULAR GLAND PROTEIN PRECURSOR.
 GN SMGB1/SMGB2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SUBMANDIBULAR GLAND;
 RX MEDLINE=92129360; PubMed=1370829;
 RA Mirels L., Ball W.D.;
 RT "Neonatal rat submandibular gland protein SMG-A and parotid secretory
 RT protein are alternatively regulated members of a salivary protein
 RT multigene family.";
 RL J. Biol. Chem. 267:2679-2687(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SUBMANDIBULAR GLAND;
 RX MEDLINE=98129760;
 RA Mirels L., Miranda A.J., Ball W.D.;
 RT "Characterization of the rat salivary-gland B1-immunoreactive
 RT proteins";
 RL Biochem. J. 330:437-444(1998).
 DR EMBL: M83210; AAC12783.1; -
 KW Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 206
 FT "NEONATAL SUBMANDIBULAR GLAND PROACINAR

Db 132 LELSVKLDVTAEVYAVRDSYGRSRL-VIGDCIYPPGSLRISLLNRLGPLQN-LIDSULTDI 188

Qy 170 LKSTVSSLLQKEICPLIRIFHISLDVNVYIQQVD 203

Db 190 LTRVIFGLVGGVVCPLVNGVLSLIDVTLAHVDAD 223

RESULT 9

Q9D794 PRELIMINARY; PNT: 270 AA.

AC Q9D794;

DT 01-JUN-2001 (TRENBLrel. 17, Created)

DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)

DE 2310021H06RIK PROTEIN.

GN 2310021H06RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI_TaxId=10090;

RN [1]

SEQUENCE FROM N.A.

RP STRAIN=C57BL/6J; TISSUE=TONGUE;

RX KADWJ J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Kawai J., Kishikawa H., Komoto H., Adachi J., Fukuda S., Arakawa T., Hara A., Fukunishi Y., Kiyosawa H., Kondo S., Yamanaka I., Aizawa K., Izawa M., Nishii K., Kiyosawa H., Kasukawa T., Saito R., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Kasahara T., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Kocova H., Flieschmann W., Gaasterland T., Gissi C., King B., Quackenbush J., Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh M.F., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seiya T., Shibata Y., Storch K.F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming U., Wyznshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., Hayashizaki Y.

RT Functional annotation of a full-length mouse cDNA collection.;

RL Nature 409:685-690(2003).

DR EHML; AK009441; BAB26290.1; -.

DR MGD; MGI:1914385; 2310021H06RIK.

SQ SEQUENCE 270 AA; 29223 MW; DB63327561D910AA CRC64;

Query Match 14.0%; Score 148.5; Db 11; Length 270;

Best Local Similarity 25.3%; Pred. No. 0.00029;

Matches 39; Conservative 38; Mismatches 74; Indels 3; Gaps

Qy 50 LLNVTSKLLPTNTDFGLKISNLSLIDVRAEPIIDGCKGLNLSFPVTANVTAGPIIGQI 1099

Db 108 VLGVKVISS-IPLLNLIIDRVNLPQLLEIGLVQSYDFHLYVTIPLGDLRVNLTLLVGS 166

Qy 110 INKASLDLLTAVTTETDPTQTHQVAVLGCASDP;SISLLDKHSQTIKNFNSVINT 169

Db 167 LELSVKLDVTAEVYAVRDSYGRSRL-VIGDCIYPPGSLRISLLNRLGPLQN-LIDSULTDI 224

Qy 170 LKSTVSSLLQKEICPLIRIFHISLDVNVYIQQVD 203

Db 225 LTRVIFGLVGGVVCPLVNGVLSLIDVTLAHVDAD 258

RESULT 10

Q9CQX3 PRELIMINARY; PNT: 270 AA.

ID Q9CQX3

AC Q9CQX3;

DT 01-JUN-2001 (TRENBLrel. 17, Created)

DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)


```

QY 18 KGILEKLVKH-----GVLOKSSAWOLAKAKAEKLLNNVSKLLPTNTDIFGLKIS 71
DB 18 KGILEKLVKH-----GVLOKSSAWOLAKAKAEKLLNNVSKLLPTNTDIFGLKIS 71
QY 34 KVIKELQELKHKNATSILOQLPLLSAMREKPAAGIPVLGSLVNTVL---KHVINLKV 90
DB 34 KVIKELQELKHKNATSILOQLPLLSAMREKPAAGIPVLGSLVNTVL---KHVINLKV 90
QY 72 NSLLLDVKAEPIDDDGKGLNLSFPVTANTVAG---PIIQIILNKASLDLITAVTETDP 128
DB 72 NSLLLDVKAEPIDDDGKGLNLSFPVTANTVAG---PIIQIILNKASLDLITAVTETDP 128
QY 91 TANILQLOVKPSANIQELVKIPLD---MVAGFNTPLVKTVFPHMTTE---AQATIRMDT 145
DB 91 TANILQLOVKPSANIQELVKIPLD---MVAGFNTPLVKTVFPHMTTE---AQATIRMDT 145
QY 129 OTHOPV-AVLGECASDPTSIISLLDKHSQIINKFVNSVINTLSTVSSLLQKEICPLIR 187
DB 129 OTHOPV-AVLGECASDPTSIISLLDKHSQIINKFVNSVINTLSTVSSLLQKEICPLIR 187
QY 146 SASGPTRLVLSOCATSHSGSLRIQLLKLHSLFLNALAKQVMNLLVPSPLNVLKNCQCPVIE 205
DB 146 SASGPTRLVLSOCATSHSGSLRIQLLKLHSLFLNALAKQVMNLLVPSPLNVLKNCQCPVIE 205
QY 188 IFHSLDVNVICQV 201
DB 188 IFHSLDVNVICQV 201
QY 206 ASFGNGYADLLGLV 219
DB 206 ASFGNGYADLLGLV 219

RESULT 13
Q61114
ID Q61114 PRELIMINARY: PRT: 474 AA.
AC Q61114
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE VON EBNER MINOR SALIVARY GLAND PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS-WEBSTER;
RA Shead M.L., Villanueva J., Paine M.L., Lei Y.P., Zhu D.H., Lusis J.,
RA Xia Y.-R., Yang J.-N.;
RA Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U46068; AAA87581.2; -
DR InterPro; IPR001124; LBP-BP1-CETP.
DR SMART; SM00328; BP1; 1.
DR SMART; SM00329; BP12; 1.
DR SMART; SM00329; BP12; 1.
SQ SEQUENCE 474 AA; 52373 MW; 377DBFA5E736709B CHC64;

Query Match 12.2%; Score 129; DB 11; Length 474;
Best Local Similarity 21.8%; Pred. No. 0.022;
Matches 42; Conservative 49; Mismatches 76; Indels 26; Gaps 6;

QY 5 VLHGLETV-----DNTIKGILEKLVGLVLOKSSAWOLAKAKAEKLLNNV 55
DB 5 VLHGLETV-----DNTIKGILEKLVGLVLOKSSAWOLAKAKAEKLLNNV 55
QY 28 VNLGPEVTKHILTKAKKHDATAILOELPL-LKAMQKSC-----SIPILDSFV 76
DB 28 VNLGPEVTKHILTKAKKHDATAILOELPL-LKAMQKSC-----SIPILDSFV 76
QY 56 SKLLPTNTDIFGLKISNLSLIDVKAEPIDDDGKGLNLSFPVTANTVAGPIIQIILNKAS 115
DB 56 SKLLPTNTDIFGLKISNLSLIDVKAEPIDDDGKGLNLSFPVTANTVAGPIIQIILNKAS 115
QY 77 HTVLKY---LWMKVTANILQDQVSTYNOELVVRIPLOWVAGINTPLIKTIVEFOMS 133
DB 77 HTVLKY---LWMKVTANILQDQVSTYNOELVVRIPLOWVAGINTPLIKTIVEFOMS 133
QY 116 LDLLTAVTETDPTQTHOPVAV-LGECASDPTSIISLLDKHSQIINKFVNSVINTLSTV 174
DB 116 LDLLTAVTETDPTQTHOPVAV-LGECASDPTSIISLLDKHSQIINKFVNSVINTLSTV 174
QY 134 TEVQALIRVERSKS--GPAHLNLSUCSSNESTIRLSILHLKLSFVVNSLAKNMNLLVPAL 191
DB 134 TEVQALIRVERSKS--GPAHLNLSUCSSNESTIRLSILHLKLSFVVNSLAKNMNLLVPAL 191
QY 175 SSLLQKEICPLIR 187
DB 175 SSLLQKEICPLIR 187
QY 192 PQIVKKNHLCPMIQ 204
DB 192 PQIVKKNHLCPMIQ 204

RESULT 14
Q909J8
ID Q909J8 PRELIMINARY: PRT: 232 AA.
AC Q909J8
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 1700058C13RIK.
GN 1700058C13RIK.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa T., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gliss C., King B., Kochava H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RA "functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
DR EMBL; AK006829; BAB24760.1; -.
DR MGD; MGI:1920638; 1700058C13RIK.
DR MGD; MGI:1920638; 1700058C13RIK.
SQ SEQUENCE 232 AA; 25713 MW; 0D52D24A3076D5DC CRC64;

Query Match 11.9%; Score 126; DB 11; Length 232;
Best Local Similarity 21.2%; Pred. No. 0.016;
Matches 41; Conservative 43; Mismatches 73; Indels 36; Gaps 6;

QY 20 ILEKLVGLVGLVLOKSSAWOLAKAKAEKLLNNVSKLLPTNTDIFGLKISNLSLIDVK 79
DB 20 ILEKLVGLVGLVLOKSSAWOLAKAKAEKLLNNVSKLLPTNTDIFGLKISNLSLIDVK 79
QY 62 LLDRNLVSGTVAPGMVGLGCMNTQQQEISINI-----TNVL----- 101
DB 62 LLDRNLVSGTVAPGMVGLGCMNTQQQEISINI-----TNVL----- 101
QY 80 AEPIDCGKGLNLSFP---VTANVTAG-----PIIQIILNKASLDLITAVTETDPTQ 130
DB 80 AEPIDCGKGLNLSFP---VTANVTAG-----PIIQIILNKASLDLITAVTETDPTQ 130
QY 102 -----DCGGIQMAFPKEWFSANITLFDIEFKLPFNLSIIKTHACMGTAESWLEKDEFG 156
DB 102 -----DCGGIQMAFPKEWFSANITLFDIEFKLPFNLSIIKTHACMGTAESWLEKDEFG 156
QY 131 HOPVAVLGECASDPTSIISLSL-LDKHSQIINKFVNSVINTLSTVSSLLQKEICPLIRIF 189
DB 131 HOPVAVLGECASDPTSIISLSL-LDKHSQIINKFVNSVINTLSTVSSLLQKEICPLIRIF 189
QY 157 KREL-VMGRCRMEPSSGSGASMSSTETSPKMKHFLHNLRESLCKVIPNLVESQVCPICEI 215
DB 157 KREL-VMGRCRMEPSSGSGASMSSTETSPKMKHFLHNLRESLCKVIPNLVESQVCPICEI 215
QY 190 IHSILDVNVICQV 202
DB 190 IHSILDVNVICQV 202
QY 216 LROLDVKLLKGLV 228
DB 216 LROLDVKLLKGLV 228

RESULT 15
Q9BOP8
ID Q9BOP8 PRELIMINARY: PRT: 199 AA.
AC Q9BOP8
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BA49G10.6 (SIMILAR TO MURINE VON EBNER MINOR SALIVARY GLAND PROTEIN,
DE ISOFORM 1) (FRAGMENT).
GN DJ118734.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RA Tracey A.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL121901; CAC03550.1; -.

```

Wed Aug 7 05:47:02 2002

```
FT NON_TER 199 199
SQ SEQUENCE 199 AA: 21517 MW: 49A4CC2143BE04B1 CRC64;

Query Match      10.8%; Score 115; DB 4; Length 199;
Best Local Similarity 24.1%; Pred. No. 0.098;
Matches 42; Conservative 38; Mismatches 76; Indels 18; Gaps 6;

QY 18 KGILEKIKVPL-----GVLOKSSAWOLAKQAEKLNNVISKLLPTNTDIEGLKIS 71
Db 34 KVIKERLTQELKDHNATSILOQLPILLSAMREKXPAGGIPVLSLVNTVL---KHIIWLKVI 90
QY 72 NSLILDVKAEPIDDGKGLNLSFPVTANVTAG---PIIGCIINLKASLDLLTAVTETDTP 128
Db 91 TANILQLOVRKPSANDQELLVKIPLD---MVAGENTPLVKTIIVEFHMTTE--AQATIRMDT 145
QY 129 QTHOPV-AVIGEGCASDPTSTLSLDKHSOIIINKFVNSVINTLKSTVSSLUKE 181
Db 146 SASGPTRELVLSDCATSHGSLRIQLLHKLFLYNALAKQVMNLLVPSPNLVKNQ 199
```

Search completed: August 6, 2002, 17:09:29
Job time: 931 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 16:52:52 ; Search time 138.55 Seconds
(without alignments)
172.363 Million cell updates/sec

Title: US-10-020-139-2_COPY_35_249
Perfect score: 1061
Sequence: 1 KLEPVLHLEGLETVDNLTGK.....NVIQQVDNPOKHQTLQTLI 215

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_032802.*
1: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/cgcdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1061	100.0	249	19 AAW69221	Human parotid secr
2	1061	100.0	249	19 AAW60682	Human parotid secr
3	1061	100.0	249	21 AAB24069	Human PRO1025 prot
4	1061	100.0	249	21 AAB25765	Human secreted pro
5	1061	100.0	249	22 AAB75351	Human secreted pro
6	1023	96.4	260	22 AAM25745	Human protein sequ
7	246	23.2	50	22 ABB41435	Peptide #8941 enco
8	246	23.2	50	22 AAM62308	Human brain expres
9	246	23.2	50	22 AAW75111	Human bone marrow
10	246	23.2	50	22 AAM35227	Peptide #9264 enco
11	182.5	17.2	278	22 AAE05367	Mouse 28.6 kDa sec

12	178	16.8	256	20 AAY06408	Human secreted pro
13	178	16.8	256	20 AAW95463	LS170 polypeptide
14	178	16.8	256	21 AAY69164	Amino acid sequenc
15	178	16.8	256	22 AAY39019	Human secreted pro
16	178	16.8	256	22 AAM29210	Human PRO polypept
17	178	16.8	256	22 AAM39721	Human polypeptide
18	178	16.8	256	22 AAB97366	Human LUNX protein
19	178	16.8	256	22 AAM92209	Human digestive sy
20	178	16.8	264	22 AAM41507	Human polypeptide
21	168	15.8	187	22 AAM92212	Human digestive sy
22	167	15.7	191	22 AAM92214	Human digestive sy
23	146	13.8	318	22 AAM47220	Human NOV6a protei
24	146	13.8	320	22 AAM47214	Human NOV6a protei
25	146	13.8	484	21 AAY77126	Human neurotransmi
26	145	13.7	484	22 AAG53976	Amino acid sequenc
27	144	13.6	484	21 AAY93375	Human PRO1357 (UNQ
28	144	13.6	484	22 AAM29163	Human PRO polypept
29	144	13.6	484	22 AAB87564	Human PRO1357. Ho
30	144	13.6	484	22 AAB66124	Protein of the inv
31	143	13.5	565	22 AAG08520	Novel human diagno
32	108	10.2	197	21 AAY86219	Human secreted pro
33	108	10.2	221	21 AAB58378	Lung cancer associ
34	98.5	9.3	1116	15 AAR60608	Tobamovirus repli
35	95	9.0	679	22 AAM35585	Haemophilus influe
36	94.5	8.9	295	22 AAU34151	Staphylococcus aur
37	94.5	8.9	299	22 AAU37067	Staphylococcus aur
38	94.5	8.9	751	22 AAU34336	Staphylococcus aur
39	94.5	8.9	825	22 AAU37370	Staphylococcus aur
40	94	8.9	298	22 AAB68518	Human novel cytoki
41	92	8.7	2383	21 AAB15945	E. coli proliferat
42	91	8.6	2710	17 AAR95016	C. difficile toxin
43	91	8.6	2710	19 AAM68387	Clostridium diffic
44	90	8.5	2478	22 AAB64182	Drosophila melanog
45	89	8.4	1057	22 AAG67419	Amino acid sequenc

ALIGNMENTS

RESULT 1
AAW69221
ID AAW69221 standard; Protein: 249 AA.
XX AC AAW69221;
XX 16-OCT-1998 (first entry)
XX Human parotid secretory protein.
XX Parotid secretory protein.
XX Parotid secretory protein; hPSP: digestive disorder; endocrine disorder;
KW non-immune defensive disorder; immune system disorder; cancer; human;
KW therapy; diagnosis.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Peptide 1..18
FT Protein /note="signal peptide"
FT FT 19..249
FT FT /note="mature hPSP"
XX PN W09828420-AL.
XX PD 02-JUL-1998.
XX PF 18-DEC-1997; 97WO-US23522.
XX PR 23-DEC-1996; 96US-0034429.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Duan R, Ruben SM;
XX

```

XX 14-NOV-1996; 96US-0749288.
XX (JNCY-) JNCYTE PHARM INC.
XX Bandman O, Goli SK;
XX WPI: 1998-297933/26.
XX N-PSDB: AAV37699.
DR DR
XX New parotid secretory protein - useful for, e.g. treatment of cancer
XX and auto-immune disease, particularly of secretory or
PT gastrointestinal tissues
PT
XX
XX Claim 1: Fig 1A-C; 65pp; English.
XX
XX This represents a human parotid secretory protein (HPSP). Antagonists
CC that bind specifically to, and modulate activity of HPSP are used to
CC treat cancer and autoimmune diseases particularly of secretory or
CC gastrointestinal tissue, e.g. cancer of salivary gland, thyroid,
CC prostate, breast, gastrointestinal tract or pancreas, Sjorgen's syndrome,
CC Graves disease, thyroiditis, insulin-dependent diabetes, pancreatitis,
CC ulcerative colitis, Crohn's disease and atrophic gastritis. Host cells
CC containing expression vectors comprising the HPSP nucleic acid are used
CC to produce recombinant HPSP which is used to generate antibodies and to
CC screen for its antagonists. Antibodies are useful directly as
CC antagonists, to transport drugs to HPSP-expressing cells, to detect cells
CC that express HPSP, to monitor patients being treated with HPSP, and for
CC purification of HPSP from natural sources. Expression of HPSP may
CC indicate cell proliferation. HPSP nucleic acid or its fragments are used
CC to detect HPSP-encoding sequences (optionally after amplification by PCR
CC by hybridisation, particularly for diagnosis and monitoring of disease,
CC but also for mapping the chromosomal sequence.
XX
XX Sequence 249 AA;
XX
Query Match 100.0%; Score 1061; DB 19; Length 249;
Best Local Similarity 100.0%; Pred. NO. 7.7e-93;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 1 KLEPVLHGEIVDNTLKIGLEKLVGLVQKGSAAWLAKQRAQAEKLLNNVISKLLP 60
DB 35 kLEPVlhgeIvDnTLkGleKlVglVqKGSaawLAKqRAQAEKllNNvISkLLp 94
QY 61 TNPDIPLGLKISNLSILDVKAEPIDGKGLNLSFPVTANVTAVGPIIQIINLKASLDLLT 120
DB 95 tntdIplgKlSnslIldvKaepIdgKglNlSfpvtanvtvAgpiIqgInlKasldlLt 154
QY 121 AVTIETDPOTHOPVAVLGGCASDPTSLSLDLKHQSIOINKFVNSVINTLKSTVSSLLQK 180
DB 155 avtIetDpThOpvAvlGgCaSDpTslSlDLKhqSIoInKfVnSVIntlKstVssllQk 214
QY 181 EICPLIRIFHSLDVNNVIOQVVDFNFOHKTQTQLTI 215
DB 215 eICPlIrIfhSlDvNnVIOqVvDfNfOHktQtqlI 249
RESULT 3
AAB24069
ID AAB24069 standard; Protein; 249 AA.
XX AAB24069;
XX
XX 29-JAN-2001 (first entry)
XX
XX Human PRO1025 protein sequence SEQ ID NO:38.
XX
XX Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
KW proliferation; tumorigenesis; identification; cancer; cytostatic;
KW neurotropic; neuroprotective; antiinflammatory; immunosuppressive;
KW immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
KW neuronal disorder; glial disorder; astrocytal disorder; angiogenic;

```

KW	hypothalamic disorder; glandular disorder; macrophagal disorder;
KW	epithelial disorder; stromal disorder; blastocoelec disorder;
KW	inflammatory disorder; immunologic disorder.
OS	Homo sapiens.
PN	WO200053755-A2.
XX	14-SEP-2000.
PD	06-JAN-2000; 2000WO-US00376.
XX	08-MAR-1999; 99WO-US05028.
PR	02-JUN-1999; 99WO-US12252.
PR	23-JUN-1999; 99US-0141037.
PR	07-JUL-1999; 99US-0143048.
PR	26-JUL-1999; 99US-0145698.
PR	30-NOV-1999; 99WO-US28313.
PR	20-DEC-1999; 99WO-US30911.
PR	05-JAN-2000; 2000WO-US00219.
XX	(GETH) GENENTECH INC.
PA	Ashkenazi AJ, Haker KP, Goddard A, Gurney AL, Willan KJ, Roy MA;
PI	Watanabe CK, Wood WI;
XX	WPI: 2000-572270/53.
DR	N-PSDB; AAC58379.
DR	Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
PT	treatment, diagnosis and prevention of cancer -
PT	Claim 61; Fig 26; 286pp; English.
XX	The present invention describes an isolated antibody that binds to
XX	one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
CC	PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
CC	PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
CC	PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
CC	PRO32145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
CC	growth. The PRO polypeptides and nucleotides are useful in the
CC	treatment, diagnosis and prevention of cancer. The antibodies and other
CC	anti-tumour compounds may be used to treat various conditions, including
CC	those characterised by overexpression and/or activation of the amplified
CC	PRO genes. Exemplary conditions or disorders to be treated with such
CC	antibodies and other compounds include benign or malignant tumours
CC	(e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC	colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
CC	carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
CC	leukaemias and lymphoid malignancies, and other disorders such as neuronal,
CC	gliat, astrocytal, hypothalamic and other glandular, macrophagal,
CC	epithelial, stromal and blastocoelec disorders, and inflammatory,
CC	angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
CC	primers and hybridisation probes used in the isolation of the human PRO
CC	sequences. AAC58367 to AAC58396 and AAC24057 to AAC24089 represent human
CC	PRO polynucleotide and protein sequences given in the exemplification of
CC	the present invention.
XX	Sequence 249 AA;
XX	Query Match 100.0%; Score 1061; DB 21; Length 249;
XX	Best Local Similarity 100.0%; Pred. No. 7.7e-93;
XX	Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 KLEPVLHLEGLTVDNTLKGILEKLVLDGLVQLKSSAWQLAKQKQAQAEKLLNNVTSKLLP 60
DB	35 klepvlhleglervdntlkgilekldvlgvlqksavqlakqkaeakllnnvtskllp 94
QY	61 TWTDTGLKISNLSILLDVKAEPIDGKGLNLVSFPVTANVTVAQPIICQIINUKASLDLTL 120
DB	95 tntdtfgklnslldvkaepidgkglnlvsfvtanvtvavpigliqglnkasldlilt 154

Query Match 100.0%; Score 1061; DB 21; Length 249;
 Best Local Similarity 100.0%; Pred. No. 7.7e-93;
 Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLEPVLHGLETVNTLKGILEKLVQKSSAWQLAKQAEKLLNNVISKLLP 60
 Db 35 klepvlhgletlvntlkgileklvdlgvlqkssawqlakqaeakllnnviskllp 94
 Qy 61 TNDIFGLKLSNLSLLDVKAEPIDDDGKGLNLSFPTANTVTAGPIIQIINLKASLDILT 120
 Db 95 tndifglklsnlsllldvkaepiddgkglnlsfpvtanvtvagiipiqiinkasldilt 154
 Qy 121 AVTETDPQTHQPVAVLGECASTPSISLSLLDKHSQIINKFVNSVINLKSTVSSLLQK 180
 Db 155 avtietdpqthqpavvlgecasdptsislsldkhsqilnkfvnsvintlkstsvsllqk 214
 Qy 181 EICPLIRIFHSLDVNVVIOQVVDNPNQHKTOLOTLI 215
 Db 215 eicplirifhslldvnnvqqvvdnpqhkgtqltli 249

RESULT 5
 AAB75351
 ID AAB75351 standard; protein: 249 AA.

AC AAB75351;
 DT 05-APR-2001 (first entry)
 DE Human secreted protein #10.

Secreted protein; prevention; treatment; diagnosis; disease;
 infection.

OS Homo sapiens.
 PN WO200100806-A2.
 PD 04-JAN-2001.

PF 21-JUN-2000; 2000WO-1B00951.
 PR 25-JUN-1999; 99US-0141032.
 PR 21-DEC-1999; 99US-0469099.
 XX (GEST) GENSET.

PI Dumas Milne Edwards J, Rouquelolet L, Jobert S;
 DR WPI; 2001-071487/08.

XX 49 Secreted proteins and the nucleic acids encoding them, useful in
 PT gene therapy and for detecting similar sequences in samples .

PS Claim 10; Page 281; 307pp; English.

XX The present invention relates to 49 Secreted proteins and the cDNAs
 CC encoding them. The protein and nucleic acids may be used in the
 CC prevention, treatment and diagnosis of diseases associated with
 CC inappropriate protein expression.

XX Sequence 249 AA;

Query Match 100.0%; Score 1061; DB 22; Length 249;
 Best Local Similarity 100.0%; Pred. No. 7.7e-93;
 Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLEPVLHGLETVNTLKGILEKLVQKSSAWQLAKQAEKLLNNVISKLLP 60
 Db 35 klepvlhgletlvntlkgileklvdlgvlqkssawqlakqaeakllnnviskllp 94

Qy 61 TNDIFGLKLSNLSLLDVKAEPIDDDGKGLNLSFPTANTVTAGPIIQIINLKASLDILT 120
 Db 95 tndifglklsnlsllldvkaepiddgkglnlsfpvtanvtvagiipiqiinkasldilt 154
 Qy 121 AVTETDPQTHQPVAVLGECASTPSISLSLLDKHSQIINKFVNSVINLKSTVSSLLQK 180
 Db 155 avtietdpqthqpavvlgecasdptsislsldkhsqilnkfvnsvintlkstsvsllqk 214
 Qy 181 EICPLIRIFHSLDVNVVIOQVVDNPNQHKTOLOTLI 215
 Db 215 eicplirifhslldvnnvqqvvdnpqhkgtqltli 249

RESULT 6
 AAM25745
 ID AAM25745 standard; Protein: 260 AA.

AC AAM25745;
 DT 16-OCT-2001 (first entry)
 DE Human protein sequence SEQ ID NO:1260.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiac; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiagregant; haemostatic; vulnery; antidiabetic; cytostatic;
 KW dermatological; antiallergic; antiasthmatic; antiparkinsonian; infection;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; inflammation;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.

XX Homo sapiens.
 OS WO200153455-A2.
 PN 26-JUL-2001.
 PD 22-DEC-2000; 2000WO-US35017.
 PF 23-DEC-1999; 99US-0471275.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Drmanac RT;
 XX WPI; 2001-457603/49.
 PI N-PSDB; AAH99686.

XX Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection .
 XX Claim 20; Page 260; 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnery;
 CC antidiabetic; cytostatic; dermatological; antiallergic; antiasthmatic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine

CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.

XX Sequence 260 AA:

Query Match 96.4%; Score 1023; DB 22; Length 260;

Best Local Similarity 96.7%; Pred. No. 3.4e-89;

Matches 208; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KLEPVLHEGLETVDTNTLKGILEKLVKVDLGVLOKSSAWOLAKQAQPAEKLNNVSKILP 60
 Db 46 KLEPVLHEGLETVDTNTLKGILEKLVKVDLGVLOKSSAWOLAKQAQPAEKLNNVSKILP 105
 Qy 61 TNDIFGLKISNSLIDVKAEPIDDCGKLNLSFPVTANVTAGPIIGQIINLKASLDLIT 120
 Db 106 TNDIFGLKISNSLIDVKAEPIDDCGKLNLSFPVTANVTAGPIIGQIINLKASLDLIT 165
 Qy 121 AVTIETDPQTHQPVAVLGECASTPTISLSLIDKHSQIINKEVNSVINTLKSTVSSLQK 180
 Db 166 AVTIETDPQTHQPVAVLGECASTPTISLSLIDKHSQIINKEVNSVINTLKSTVSSLQK 225
 Qy 181 EICPLIRIFIRFINSIDVNVVIOOVVDPNPOHKTOLTLI 215
 Db 226 EICPLIRIFIRFINSIDVNVVIOOVVDPNPOHKTOLTLI 260

RESULT 7

ID ABB41435

XX ABB41435 standard; Peptide: 50 AA.

AC ABB41435;

XX 04-FEB-2002 (first entry)

XX Peptide #8941 encoded by human foetal liver single exon probe.

XX Human: foetal liver; gene expression: single exon nucleic acid probe.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US006569.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human foetal liver.

XX

PS Claim 27; SEQ ID NO 34070; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at http://wipo.int/pub/published_pcl_sequences.

XX Sequence 50 AA:

Query Match 23.2%; Score 246; DB 22; Length 50;

Best Local Similarity 100.0%; Pred. No. 2.8e-16;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 104 PIIQIINLKASLDLITAVTIETDPQTHQPVAVLGECASTPTISLSLID 153
 Db 1 PIIQIINLKASLDLITAVTIETDPQTHQPVAVLGECASTPTISLSLID 50

RESULT 8

AA62308

ID AAM62308 standard; Protein: 50 AA.

XX AAM62308;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 34413.

XX Human; brain expressed exon; gene expression analysis; probe;

XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 XX epilepsy; cancer.

XX Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00657.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human
 PT brains.

XX Example 4; SEQ ID NO: 34413; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC brain. They can be used to measure gene expression in brain cell samples,

CC which may enable the diagnosis and improved treatment of nervous system

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.

SQ Sequence 50 AA:

Query Match 23.2%; Score 246; DB 22; Length 50;
 Best Local Similarity 100.0%; Pred. No. 2.8e-16;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 PIIGQIINKASLLOLTAVTETDPTQHPVAVLGECA SDPTSTLSLSD 153
 |||||
 Db 1 pligqilnkasldltavtietdptqhpqvavlgcasdptsislld 50

RESULT 9

AAM75111
 ID AAM75111 standard; Protein: 50 AA.

XX AC AAM75111;

DT 06-NOV-2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 35417.
 DE Human; bone marrow expressed exon; gene expression analysis: probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma.
 KW
 XX Homo sapiens.
 OS
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00668.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 PD WPI; 2001-488900/53.
 XX
 PF Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 PS Example 4; SEQ ID NO: 35417; 658pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.
 XX
 SQ Sequence 50 AA:

SQ Sequence 50 AA:

Query Match 23.2%; Score 246; DB 22; Length 50;
 Best Local Similarity 100.0%; Pred. No. 2.8e-16;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 PIIGQIINKASLLOLTAVTETDPTQHPVAVLGECA SDPTSTLSLSD 153
 |||||
 Db 1 pligqilnkasldltavtietdptqhpqvavlgcasdptsislld 50

RESULT 10

AAM35227

ID XX AAM35227 standard; Protein: 50 AA.

XX AC AAM35227;

DT 17-OCT-2001 (first entry)

DE Peptide #9264 encoded by probe for measuring placental gene expression.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 PD WPI; 2001-48897/53.
 XX
 PF Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 27; SEQ ID NO 35496; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP;
 CC see AAL1315-AA15746). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX
 SQ Sequence 50 AA:

Query Match 23.2%; Score 246; DB 22; Length 50;
 Best Local Similarity 100.0%; Pred. No. 2.8e-16;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 PIIGQIINKASLLOLTAVTETDPTQHPVAVLGECA SDPTSTLSLSD 153
 |||||
 Db 1 pligqilnkasldltavtietdptqhpqvavlgcasdptsislld 50

RESULT 11

AAE05367

ID AAE05367 standard; Protein: 278 AA.

XX AC AAE05367;

DT 12-SEP-2001 (first entry)

DE Mouse 28.6 kDa secreted protein.

XX Mouse; cytostatic; antitumour; immunoregulatory; tissue integrity;
 KW wound healing; immune response; vaccine; cancer; asthma; allergy;
 KW cell trafficking; therapy; 28.6 kDa secreted protein.
 XX
 OS Mus sp.
 XX

Db 65 lgiilenplldililp99ggtsggllgllgkvsvipulnnidilvtdpqllelgivqsp 124
QY 85 DGKGINLSPFVTANVTYAGPTIG-QIINLKASLDLLTAVTIETDPQTHQPVAVLGCASD 143
Db 125 dghrlyvctipiglkivntplvqasllrlavklditaeilavrdkqerihl-vlgdcths 183
QY 144 PTSLSLSDKHSQI-INKFVNSVINTLKSTVSSLQKEICPLIRIFIHSLDNNVIOQV 202
Db 184 pgsqisilldglplpqlgldstgilnkvlpelvqgnvcplvnevlrglditlvhdv 243
QY 203 DNPQHKQ 210
Db 244 nmllhgiq 251

RESULT 15
AAU35019
ID AAU39019 standard; Protein: 256 AA.
AC AAU39019;
XX
DT 16-JAN-2002 (first entry)
XX
DE Human secreted protein nh796_1.
XX
KW Human; secreted protein; antiinflammatory; immunosuppressive;
KW neurotropic; neuroprotective; antiarthritic; antimicrobial; vulnery;
KW cytosolic; antidiabetic; virucide; antifertility; anticonvulsant;
KW vasotropic; antiparkinsonian; immunostimulant; dermatological;
KW antirheumatic; antitumor; antilucer; osteopathic; tranquiliser;
KW cerebroprotective; cytokine; cell proliferation; cell differentiation;
KW immune deficiency; severe combined immunodeficiency; SCID; tumour;
KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;
KW graft-versus-host disease; myeloid deficiency; wound healing; ulcer;
KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; infection; cardiac disease;
KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;
KW food supplement; vaccine.
XX
OS Homo sapiens.
XX
PN WO200175068-A2.
XX
PD 11-OCT-2001.
XX
PF 22-MAR-2001; 2001w0-US09369.
XX
PR 30-MAR-2000; 2000US-0539330.
PR 04-DEC-2000; 2000US-0729674.
XX
PA (GEM) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;
PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;
PI Clark H, Fechtel K, Merberg D;
XX
DR WPI; 2001-639363/73.
DR N-PSDB; AAS59237.
XX
PT Secreted human proteins, useful as vaccine for treating various
PT diseases such as autoimmune disorders (e.g. multiple sclerosis), and
PT nervous system disorders (e.g. stroke).
XX
PS Disclosure; Page 497-498; 619pp; English.
XX
CC The invention relates to novel human secreted proteins, the nucleic
CC acids encoding them. The protein may exhibit cytokine, cell proliferation
CC or cell differentiation activity or may induce production of other
CC cytokines in certain cell populations and may exhibit immune stimulating
CC or immune suppressing activity, which is useful for the treatment of
CC various immune deficiencies and disorders e.g. severe combined
CC immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,

CC systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
CC inflammation. The proteins are also useful in the treatment of diseases
CC and disorders including tissue, skin and organ transplantation and in
CC graft-versus-host diseases (GVHD), in the induction of tumour immunity,
CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,
CC in the treatment of burns, incisions and ulcers; as well as in treatment
CC of periodontal diseases, osteoporosis or osteoarthritis, mediated by
CC inflammatory processes, diseases of the peripheral nervous system,
CC Alzheimer's, Parkinson's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, and Shy-Drager syndrome. Infections,
CC infarction of cardiac and central nervous system vessel e.g. stroke,
CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The
CC protein, having activin- or inhibin-related activities is useful as a
CC contraceptive based on the ability of inhibins to decrease fertility in
CC female mammals and decrease spermatogenesis in male mammals. The
CC proteins and nucleic acids are also useful as food supplements. The
CC present sequence represents a secreted protein of the invention.
XX
SQ Sequence 256 AA;
Query Match 16.8%; Score 178; DB 22; Length 256;
Best Local Similarity 24.5%; Pred No 7.5e-09;
Matches 46; Conservative 44; Mismatches 92; Indels 6; Gaps 4;
QY 28 LGVLOKSSAWOLAKOAEKLLNNVISK---LLPTMTDIFGLKINSLLDVKAEPID 84
Db 65 lgiilenplldililp99ggtsggllgllgkvsvipulnnidilvtdpqllelgivqsp 124
QY 85 DGKGINLSPFVTANVTYAGPTIG-QIINLKASLDLLTAVTIETDPQTHQPVAVLGCASD 143
Db 125 dghrlyvctipiglkivntplvqasllrlavklditaeilavrdkqerihl-vlgdcths 183
QY 144 PTSLSLSDKHSQI-INKFVNSVINTLKSTVSSLQKEICPLIRIFIHSLDNNVIOQV 202
Db 184 pgsqisilldglplpqlgldstgilnkvlpelvqgnvcplvnevlrglditlvhdv 243
QY 203 DNPQHKQ 210
Db 244 nmllhgiq 251

Search completed: August 6, 2002, 16:52:52
Job time: 334 sec

Matches 65: Conservative 32: Mismatches 99: Indels 52: Gaps 13:

```

QY 2 LEPVLHREGIETVONTLK--GILEKIKVDVLCV-----LQSSAWQLAKOQAEAK 49
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 80 LQPI--GENLETIRNQLPTRRRRFACVIGLAA--LCVATAAQVTAVALVKANKNAAAI 137
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 50 L--LNKVISKLPTNTDIFGLKISNLSLDVKAEPIDGKGLNLSFPVT--ANVTVAGPIIG 107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 LNKIKAIQK---TNTAVADV--VQATQSLGTAVQAVQVQIINSVSPAITAANCKAQDAIIG 193
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 108 QLINIKASLDLITAVTITETDPQTHOPVAVILGECASDPTSIS--LSLLDKHSOIKNFVN 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 SILNL-----YLTHTITTFINQITNP-----ALSPITIQALRILLGSLPTVVEKSN 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 165 SVNTLKTSVSSLLQKEICPI-----IRIFHSLDVNVVIOQVVD-----NPQH 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 TQISAELLSSGILITGGQIVGHDLTYMQVMQIKIELPTLTQVQATQIIDLATISAFINQEV 301
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 208 KTQLQTLI 215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 MAQLPTRV 309
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 2
US-08-480-604A-6

; Sequence 6, Application US/08480604A
; Patent No. 5736139

; GENERAL INFORMATION:

; APPLICANT: KINK, JOHN A.

; APPLICANT: THALLEY, BRUCE S.

; APPLICANT: PADHYE, NISHA V.

; APPLICANT: FIRCA, JOSEPH R.

; APPLICANT: STAFFORD, DOUGLAS C.

; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND

; PREVENTION OF C. DIFFICILE DISEASE

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MEDLEN & CARROLL, LLP

; STREET: 220 MONTGOMERY STREET, SUITE 2200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/480,604A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/422,711

; FILING DATE: 14-APR-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/405,496

; FILING DATE: 16-MAR-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/329,154

; FILING DATE: 25-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/161,907

; FILING DATE: 02-DEC-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/985,321

; FILING DATE: 04-DEC-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/429,791

; FILING DATE: 31-OCT-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: INGOLIA, DIANE E.

; REGISTRATION NUMBER: 40,027

; REFERENCE/DOCKET NUMBER: OPHU-01763

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 705-8410

; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2710 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-480-604A-6

Query Match 8.6%; Score 91; DB 1; Length 2710;

Best Local Similarity 20.7%; Pred. No. 3.8;

Matches 50; Conservative 51; Mismatches 98; Indels 42; Gaps 10;

```

QY 1 KLEPVLHREGIETVONTLK--GILEKIKVDVLCVLOK-----SSAWQLAKOQAEAK 49
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 842 KLEPVKNIIHNSIDGLIDEFNLLNVSDELYELKKNLNDKYLISFEDISKNNSTYSVR 901
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 50 LLN--NVISKILPTNTDIFG-----LKTNSNLSLDVKAEPIDGKGLNLSFVTA 97
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 902 FINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVGNLLDN---IQLDHTSQV 958
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 98 NVTAGPIIGIQLINLKASLDLITAVTITETDPQTHOPVAVILGECASDPTSISLLDKHSQ 157
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 959 NTLNAAFFIQLSLIDYSSNKNVDLSTSVKVLQY---AQLFSTGLNTIYDSIQLVN--- 1011
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 158 IINKFVNSVINTLKRSTVSSLLQKEICPLRIFHSLDVN--VIOQVVD--NPOHKTULOTL 214
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1012 LFSNAVNDTINVLTITEGI-----PIVSTILDGNGLGAIKELLDHDPPLKKKELEAK 1065
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 215 J 215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1066 V 1066
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 3

US-08-405-496A-6

; Sequence 6, Application US/08405496A

; Patent No. 5919665

; GENERAL INFORMATION:

; APPLICANT: WILLIAMS, JAMES A.

; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM

; TITLE OF INVENTION: NEUROTOXIN

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MEDLEN & CARROLL, LLP

; STREET: 220 MONTGOMERY STREET, SUITE 2200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/405,496A

; FILING DATE: 16-MAR-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/329,154

; FILING DATE: 25-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/161,907

; FILING DATE: 02-DEC-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/985,321

; FILING DATE: 04-DEC-1992

; PRIOR APPLICATION DATA:

```
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-405-496A-6

Query Match      8.6%; Score 91; DB 2; Length 2710;
Best Local Similarity 20.7%; Pred. No. 3.8;
Matches 50; Conservative 51; Mismatches 98; Indels 42; Gaps 10;

Qy 1 KLEPVLHEGLTVDNTLK--GILEKLVOLGVLOK-----SSAWOLAKQKAEAK 49
      ||||| : : : : : || : : : : : || : : : : :
Db 842 KLEPVKNLIHNSIDDLDEFNLLENVSDLEYELKLNLDKYLISFEDISKNNSTYSVR 901
      : : : : : : : : : : : : : : : : : : : : :

Qy 50 LLN--NVISKLLPTNTDIFG-----LKISNLIIDVKAEPIDDGKGLNLSFPVTA 97
      : : : : : : : : : : : : : : : : : : : : :
Db 902 FINKSGESVYVETEKEIFSKYSEHITKEISTTKNSITDVGNNLLDN---IQLDHTSQV 958
      : : : : : : : : : : : : : : : : : : : : :

Qy 98 NTVVAGPIIGOIINLKASLDLLTAVTETDPTQHPVAVLGEACASDPTSLSLLDKHSQ 157
      | | | | : : : : : | | : : : | | : : : | | : : :
Db 959 NTLNAAFFIQSLIDYSSNKDVNDLSTSVKQVLY---AQLFSTGLNTIYDSIQLVN--- 1011
      : : : : : : : : : : : : : : : : : : : : :

Qy 158 IINKFVNSVINTLKSTVSSLLQKEICPLRIFIHSLDVN-VIOQVVD--NPQHKTOLOQL 214
      : : : : : : : : : : : : : : : : : : : : :
Db 1012 LISNAVNDTINVLPITTEGI-----PIVSTILDGINLGAAIKELLDHDPLLKKELEAK 1065
      : : : : : : : : : : : : : : : : : : : : :

Qy 215 I 215
Db 1066 V 1066

RESULT 4
US-08-915-136-6
; Sequence 6, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-915-136-6

Query Match      8.6%; Score 91; DB 4; Length 2710;
Best Local Similarity 20.7%; Pred. No. 3.8;
Matches 50; Conservative 51; Mismatches 98; Indels 42; Gaps 10;

Qy 1 KLEPVLHEGLTVDNTLK--GILEKLVOLGVLOK-----SSAWOLAKQKAEAK 49
      ||||| : : : : : || : : : : : || : : : : :
Db 842 KLEPVKNLIHNSIDDLDEFNLLENVSDLEYELKLNLDKYLISFEDISKNNSTYSVR 901
      : : : : : : : : : : : : : : : : : : : : :

Qy 50 LLN--NVISKLLPTNTDIFG-----LKISNLIIDVKAEPIDDGKGLNLSFPVTA 97
      : : : : : : : : : : : : : : : : : : : : :
Db 902 FINKSGESVYVETEKEIFSKYSEHITKEISTTKNSITDVGNNLLDN---IQLDHTSQV 958
      : : : : : : : : : : : : : : : : : : : : :

Qy 98 NTVVAGPIIGOIINLKASLDLLTAVTETDPTQHPVAVLGEACASDPTSLSLLDKHSQ 157
      | | | | : : : : : | | : : : | | : : : | | : : :
Db 959 NTLNAAFFIQSLIDYSSNKDVNDLSTSVKQVLY---AQLFSTGLNTIYDSIQLVN--- 1011
      : : : : : : : : : : : : : : : : : : : : :

Qy 158 IINKFVNSVINTLKSTVSSLLQKEICPLRIFIHSLDVN-VIOQVVD--NPQHKTOLOQL 214
      : : : : : : : : : : : : : : : : : : : : :
Db 1012 LISNAVNDTINVLPITTEGI-----PIVSTILDGINLGAAIKELLDHDPLLKKELEAK 1065
      : : : : : : : : : : : : : : : : : : : : :

Qy 215 I 215
Db 1066 V 1066

RESULT 5
US-08-790-912-4
; Sequence 4, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
```

```

: COUNTRY: USA
: ZIP: 19103-2398
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/790,912
: FILING DATE: 29-JAN-1997
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/026,939
: FILING DATE: 23-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Leary, Kathryn
: REGISTRATION NUMBER: 36,317
: REFERENCE/DOCKET NUMBER: 7600-401
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 567-2020
: TELEFAX: (215) 567-2991
: TELEX: 831-494
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1861 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-790-912-4

```

```

Query Match      8.5%; Score 90.5; DB 2; Length 1861;
Best Local Similarity 26.5%; Pred. No. 2.5;
Matches 45; Conservative 26; Mismatches 58; Indels 41; Gaps 9;

Qy 12 TVDN-TLKGLKLVGLQKSSAWQLAKOKAQAELNNVSKLLPTNTDIFGL-- 68
: ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 743 TVENLTLDVNSLCKTDIGAL-----ANEANNATRLNNVHVDVLAGERGIGLW 793

Qy 69 -----KISNSLLDVKAEPIDDCKGLMSFPVTANVTAGPIIGQIINKASLDLLTA-V 122
: ||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 794 KADNSKLSNSF-----KRIYN-SYETRAPYNIIG-LVGQTLGINALVDKSKATI 842

Qy 123 TIETD-PQTHQPVAVLGECASTDPTSLSLDDKHSHQIINKFVNSVINTLK 171
: || : || : || : || : || : || : || : || : || : || : ||
Db 843 TISSNADSTNQTVGLA-----GIVEKDAIISYABGNINNVK 881

```

```

RESULT 6
US-09-541-782-10
: Sequence 10, Application US/09541782
: Patent No. 6284480
: GENERAL INFORMATION:
: APPLICANT: Nislow, Corey
: APPLICANT: Sakowicz, Roman
: APPLICANT: Beraud, Christophe
: TITLE OF INVENTION: Antifungal Assay
: FILE REFERENCE: 1015
: CURRENT APPLICATION NUMBER: US/09/541,782
: CURRENT FILING DATE: 2000-04-03
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 10
: LENGTH: 1057
: TYPE: PRT
: ORGANISM: H.septens
: US-09-541-782-10

```

```

Query Match      8.4%; Score 89; DB 4; Length 1057;
Best Local Similarity 23.0%; Pred. No. 1.6;
Matches 53; Conservative 35; Mismatches 79; Indels 60; Gaps 11;

```

```

Qy 10 LETVDNLTGLKLVGLQKSSAWQLAKOK-----AQAEKLLNNVSKLL-----P 60
: || : || : || : || : || : || : || : || : || : || : ||
Db 447 LDQCKSLQNKTOELETQKHQETKL-QLVKEEYITSLESTEELHDAASKLLNTVEE 505
: || : || : || : || : || : || : || : || : || : || : ||
Qy 61 TNDTIFGL--KISNSLLDVKAEPIDD--GKGLNLSF-----PVTANVTAGP 104
: || : || : || : || : || : || : || : || : || : || : ||
Db 506 TTKDVSGLSKLDKKKAVDQHNAAQDIFCKNLNSLNFNNMEELIKDCSSKQKAMLEVHK 565
: || : || : || : || : || : || : || : || : || : || : ||
Qy 105 IIGQIINKAS-LDLLTAVTIETDPQTHQPVAVLGECASTDPTSLSLDDKHSHQIINKF- 162
: || : || : || : || : || : || : || : || : || : || : ||
Db 566 LFGNLLSSVSALDTITV-----ALGSLTSLPENVSTHV-----SOIFNMIL 608
: || : || : || : || : || : || : || : || : || : || : ||
Qy 163 -----VNSVINTLKSTVSSLLQKEICPLIRIFTHSLDVN 196
: || : || : || : || : || : || : || : || : || : || : ||
Db 609 KEQSLAAESKTVLQELINVLKLTOLLSSLEMLSPTV---VSILKIN 651
: || : || : || : || : || : || : || : || : || : || : ||

```

```

RESULT 7
US-08-261-660A-49
: Sequence 49, Application US/08261660A
: Patent No. 5731415
: GENERAL INFORMATION:
: APPLICANT: Gazzano-Santorio, Helene
: APPLICANT: Theofan, Georgia
: APPLICANT: Town, Patrick W.
: TITLE OF INVENTION: Lipopolysaccharide Binding Protein Derivatives
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-8402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA: US/08/261,660A
: APPLICATION NUMBER: US/08/261,660A
: FILING DATE: 17-JUN-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Sharp, Jeffrey S.
: REGISTRATION NUMBER: 31,879
: REFERENCE/DOCKET NUMBER: 27129/32137
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6300
: TELEFAX: 312/474-0448
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 49:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 198 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-261-660A-49

```

```

Query Match      7.9%; Score 83.5; DB 1; Length 198;
Best Local Similarity 18.1%; Pred. No. 0.51;
Matches 32; Conservative 44; Mismatches 76; Indels 25; Gaps 8;

```

```

Qy 44 AQAEKLLNNVSKL-LPTNTDIFGLKISNS-----LILDVK-----APIDDGKGLNLS 92
: || : || : || : || : || : || : || : || : || : || : ||
Db 18 SQGTAAALQXELKRIKIPDYSDFKIKHLGKHGHSYFSDIREFQLPSSQISWPNVGLK 77
: || : || : || : || : || : || : || : || : || : || : ||
Qy 93 FPVT-ANVTAGPIIGO--IINKASLDL-LTAVTIETD-----PQTHQPVAVLGECA 142
: || : || : || : || : || : || : || : || : || : || : ||
Db 78 FSIISNANIKISGKWKAKRFLKMSGNFDLSIEGMSISADLKLSGNPTSGKITTCSSCS 137
: || : || : || : || : || : || : || : || : || : || : ||

```


QY 131 -----H-----OPVAVIGECASDPTSLSL-----SLLDKHSQIIN 160
 DB 791 ISPKLHVDKGFVNILMTSFLQYISSEVNPFSUETDSSAPSKQLEQKGLLSLFPKPMQ 850
 QY 161 KFNVSIVNTLSTVSSLLQKEI-C-----PLIRIFHSLDVNVVIOQ 200
 DB 851 KFLDHVDL-----QVSALYAJQVHCYNSNFPKGMILLRFFVHFYDMEILIE 896

RESULT 11

US-08-790-912-3
 ; Sequence 3, Application US/08790912
 ; Patent No. 5976542
 ; GENERAL INFORMATION:
 ; APPLICANT: Weiser, Jeffrey N.
 ; APPLICANT: Plant, Andrew G.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
 ; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
 ; STREET: 1601 Market Street, 36th Floor
 ; CITY: Philadelphia
 ; STATE: Pennsylvania
 ; COUNTRY: USA
 ; ZIP: 19103-2398
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/790,912
 ; FILING DATE: 29-JAN-1997
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/026,939
 ; FILING DATE: 23-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Leary, Kathryn
 ; REGISTRATION NUMBER: 36,317
 ; REFERENCE/DOCKET NUMBER: 7600-401
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (215) 567-2020
 ; TELEFAX: (215) 567-2991
 ; TELEX: 831-494
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1964 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-790-912-3

Query Match 7.9% Score 83.5; DB 2; Length 1964;
 Best Local Similarity 24.3%; Pred. No. 14;
 Matches 49; Conservative 34; Mismatches 68; Indels 51; Gaps 10;
 QY 37 WOLAKQKQAEKLLNNVSKLLPTNTDIF-----CLKISNSL-----ILDVKAEPID 84
 DB 1004 WDRTSGEKHAGE-LTNVLSDVNVNTGNAITGYHYTKMKVANTFSSKANRVFNVTLE--K 1060
 QY 85 DGKGLNLSFPVTANVTVAGPIIGQIINKASLDLLTAVTITDPQTHQPVAVLGECA SDP 144
 DB 1061 DEVYSKSEFEEGRTMLDA-----SOIVSKKAEINPLTLPTVE-----PLSTSGKXDSDF 1109
 QY 145 TSI-----SLSLDKHSQIINKFNVSINTLKSTV-----SSLLQKEICPLIRIF 189
 DB 1110 SKIAHYQANRALVYKNTIEKLLPFYN-----KSTIVKYGNLVKNSLLYOKELLSAVMM- 1162
 QY 190 IHSLVNVIQOVVDNPOHKTQL 211

DB 1163 ---KDDQVITDIVSNKOTANKL 1181

RESULT 12

US-08-790-912-2
 ; Sequence 2, Application US/08790912
 ; Patent No. 5976542
 ; GENERAL INFORMATION:
 ; APPLICANT: Weiser, Jeffrey N.
 ; APPLICANT: Plant, Andrew G.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
 ; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
 ; STREET: 1601 Market Street, 36th Floor
 ; CITY: Philadelphia
 ; STATE: Pennsylvania
 ; COUNTRY: USA
 ; ZIP: 19103-2398
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/790,912
 ; FILING DATE: 29-JAN-1997
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/026,939
 ; FILING DATE: 23-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Leary, Kathryn
 ; REGISTRATION NUMBER: 36,317
 ; REFERENCE/DOCKET NUMBER: 7600-401
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (215) 567-2020
 ; TELEFAX: (215) 567-2991
 ; TELEX: 831-494
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2052 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-790-912-2

Query Match 7.9% Score 83.5; DB 2; Length 2052;
 Best Local Similarity 24.3%; Pred. No. 15;
 Matches 49; Conservative 34; Mismatches 68; Indels 51; Gaps 10;
 QY 37 WOLAKQKQAEKLLNNVSKLLPTNTDIF-----CLKISNSL-----ILDVKAEPID 84
 DB 1077 WDRTSGEKHAGE-LTNVLSDVNVNTGNAITGYHYTKMKVANTFSSKANRVFNVTLE--K 1133
 QY 85 DGKGLNLSFPVTANVTVAGPIIGQIINKASLDLLTAVTITDPQTHQPVAVLGECA SDP 144
 DB 1134 DEVYSKSEFEEGRTMLDA-----SOIVSKKAEINPLTLPTVE-----PLSTSGKXDSDF 1182
 QY 145 TSI-----SLSLDKHSQIINKFNVSINTLKSTV-----SSLLQKEICPLIRIF 189
 DB 1183 SKIAHYQANRALVYKNTIEKLLPFYN-----KSTIVKYGNLVKNSLLYOKELLSAVMM- 1235
 QY 190 IHSLVNVIQOVVDNPOHKTQL 211
 DB 1236 ---KDDQVITDIVSNKOTANKL 1254
 RESULT 13
 US-08-431-517F-2

```

: Sequence 2, Application US/08431517F
: Patent No. 6265187
: GENERAL INFORMATION:
: APPLICANT: Scott, Randal W
: APPLICANT: Marra, Marian N
: TITLE OF INVENTION: RECOMBINANT ENDOTOXIN-NEUTRALIZING PROTEINS
: FILE REFERENCE: 1103/113070U01
: CURRENT APPLICATION NUMBER: US/08/431,517F
: CURRENT FILING DATE: 1995-05-01
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: PatentIn ver. 2.1
: SEQ ID NO 2
: LENGTH: 481
: TYPE: PRT
: ORGANISM: human
: OTHER INFORMATION: No. 6265187e: human LBP-b DNA and amino acid sequence
US-08-431-517F-2

Query Match          7.8%; Score 83; DB 4; Length 481;
Best Local Similarity 18.5%; Pred.No. 2.1;
Matches 31; Conservative 42; Mismatches 73; Indels 22; Gaps

QY 64 DIFGLKISNLSILDVKAEPIDDCGLNLSFPYANTVTAG--PIICQIINKASLDL-LT 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 77 EPHSLMIHSCELHLSALRPV-PQGLSLIS-DSSIRVQCRKVRKSEFKLQGSFDSVK 134
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 AVTIETD-----PQTHQVAVLGECAASDPTSLSLDLKHSOIIINKFVNSVIN----TLK 171
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 135 GISISVNLILGSESSGRPTVTASSCSADIADVDMKSGDFGWLNLNFHQIESKFKVLE 194
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 172 STVSSULQKEICPLRIFHISLDV-----NVIGQVVVDNPOHKQTOL 211
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 195 SRICEMIQSVSSDLQPYLQTLPTVTEIDSFADIDYSLVEAPRAYAQM 242
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 14
US-09-004-838-121
; Sequence 121, Application US/09004838
; Patent No. 6350933
; GENERAL INFORMATION:
; APPLICANT: Michelmore, Richard W.
; APPLICANT: Shen, Kathy
; APPLICANT: Meyers, Blake
; TITLE OF INVENTION: Procedures and Materials for
; CONFERRING PEST RESISTANCE IN PLANTS
; NUMBER OF SEQUENCES: 140
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/004,838
; FILING DATE: 09-JAN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/781,734
; FILING DATE: 10-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-078810US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300

```

; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 468 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..468
; OTHER INFORMATION: /note= "RG2P deduced sequence"
US-09-004-838-121

Query Match      7.7%  Score 81.5;  DB 4;  Length 468;
Best Local Similarity 20.8%;  Pred. No. 2.9;
Matches 49;  Conservative 42;  Mismatches 90;  Indels 55;  Gaps
Gaps

QY 31 LKSSAWOLAKQKQAEKLNNNVISKLLPTNTDIFGLKISNLSILD-----VKRAEPID- 84
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 IQEAVADYLXIELKEKTSARADMLRKLKLVAKSD--GGKKKFLVLDVWQFVDELDGL 61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 85 ---DKGGLNLSPEVTAN-----VTVAGPIIGQILNKLASLD-----LLTAVTITDPPQ 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 SPLPAGQVNFKYLTSROVDVCTMMGYEANSIINMKIILLDEEAQSLFMEFVQISSDVDPK 12
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 130 THQ-----PVAV-----LGECASDPTSLSLDKHSGIINKFVNSVINTLK 17
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 LHKIGEDIVRKCGLPIAKTKMALTLENKSDAWSDALESLEHDD--LHNFVNEVFGISY 17
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 172 STVSSLLAKETCPLIRIFHSLOV-----NVIQVVDVNPQHKTQLQTL 215
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 DYLDQDETQYIFLLCGLFPEDYNIPPEELMRGYGMGNLFKKVYTIIEARARLNTCI 235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15
US-08-261-660A-51
Sequence 51, Application US/08261660A
Patent No. 5731415
GENERAL INFORMATION:
APPLICANT: Gazzano-Santoro, Helene
APPLICANT: Theofan, Georgia
APPLICANT: Town, Patrick W.
TITLE OF INVENTION: Lipopolysaccharide Binding Protein Derivatives
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum
STREET: 6100 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,660A
FILING DATE: 17-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sharp, Jeffrey S.
REGISTRATION NUMBER: 31,879
REFERENCE/DOCKET NUMBER: 27129/32137
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 197 amino acids
TYPE: amino acid

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2002, 16:01:08 ; Search time 1816.11 Seconds
(without alignments)
11845.362 Million cell updates/sec

Title: US-10-020-139-1

Perfect score: 1028
Sequence: 1 CACGAGATTTCATGAGCATC.....AAGCCCATTTCTGCAAAA 1028

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl :

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_hgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description

1	1019.8	99.2	1058	6	AX061621
2	746.8	72.6	750	9	AF432917
3	237.4	23.1	999	4	BTU79413
4	219.4	21.3	1024	4	BTU79414
5	198	19.3	974	10	BC010288
6	191	18.6	161593	9	HSBA49C10
7	190.6	18.5	867	10	MMPSPR
8	173.2	16.8	962	10	RATPSG
9	70.6	6.9	13962	10	MMPSPG
10	65.2	6.3	827	10	RATSMGB
11	62.6	6.1	210	10	MUSPSPA2
12	42.6	4.1	12334	1	AE001965
13	41	4.0	125020	9	AF429315
14	40.4	3.9	768	6	E53840
15	40.4	3.9	1011	9	AF172993
16	40.4	3.9	1015	6	E53841
17	40.4	3.9	1015	9	AB024937
18	40.4	3.9	1053	9	BC012549
19	40.4	3.9	1084	9	AF158745
20	40	3.9	227	10	MUSPSPA3
21	39	3.8	3186	9	HUMKGICC
22	38.6	3.8	837	6	AX123429
23	38.6	3.8	960	6	AX066081
24	38.6	3.8	53792	2	AC107180
25	38.6	3.8	128458	9	AL139090
26	38.6	3.8	167356	2	AC021774
27	38.6	3.8	309400	6	AC127153
28	38.4	3.7	7459	1	SC8A2
29	38.2	3.7	163908	9	AC005411
30	38	3.7	3640	9	AB056750
31	38	3.7	183991	9	AC007391
32	37.8	3.7	206499	2	AC090376
33	37.8	3.7	207767	2	AP001497
34	37.6	3.7	33330	8	AP001304
35	37	3.6	963	6	AX01906
36	37	3.6	1035	6	AX01908
37	37	3.6	1636	6	AX092346
38	37	3.6	1680	6	AX224643
39	37	3.6	1684	9	BC008429
40	37	3.6	2036	6	AX224647
41	37	3.6	105740	2	AC098322
42	36.8	3.6	306	5	D88464
43	36.8	3.6	116138	2	AC095073
44	36.8	3.6	170000	2	AC094951
45	36.6	3.6	46446	2	AC107726

ALIGNMENTS

RESULT 1

AX061621

LOCUS

DEFINITION

AX061621

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

AX061621 1058 bp DNA linear PAT 24-JAN-2001
Sequence 26 from Patent WO0100806.
AX061621 GI:12406703
human.
Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1058)
Dumas Milne Edwards, J.B., Bouqueleret, L. and Jobert, S.
Complementary DNA's encoding proteins with signal peptides
Patent: WO 0100806-A 26 04-JAN-2001;
GENSET (FR)
Location/Qualifiers
1..1058
/organism="Homo sapiens"
/db_xref="taxon:9606"
67..816
/note="unnamed protein product"
/codon_start=1

/protein_id="CAC24582.1"
 /db_xref="GI:12406704"
 /translation="MLQWLKLVLLCGVLTGTSESLDNLGNDLSNVVVKLEPVLHGL
 ETVDNTLKGLEKLVKLVGLQKSSAWQLAKQAEKLNINISKLLPTNTDIFGL
 KISNLSLDVKAEPIDDKGKLNLSFPVNTVAGPIGQIINKASLDLITAVTIEI
 DPQTHQPVAVLRCAADPTISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICP
 LLRIFIHSLDWNVIQGVVDNPQHKTLQTLI"
 67. .111
 /note="Von Heijne matrix score 5.20 seq QLWKLVLGVLGT/CT"
 polyA_signal 1023. .1028
 polyA_site 1042. .1058
 BASE COUNT 291 a 302 c 231 g 234 t
 ORIGIN

Query Match 99.2%; Score 1019.8; DB 6; Length 1058;
 Best Local Similarity 99.8%; Pred. No. 4.4e-267;
 Matches 1021; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 GATTTTCATGAGCATCTCTCTAAACGGGTGTCAGACAAAGATGCTTCAGCTTTGGAA 65
 DB 24 GATTTTCATGAGCATCTCTCTAAACGGGTGTCAGACAAAGATGCTTCAGCTTTGGAA 83
 QY 66 ACTTGTTCCTGTCGGCGCTGCTCACTGGGACCTCAGAGTCTCTCTTGCACATCTTGG 125
 DB 84 ACTTGTTCCTGTCGGCGCTGCTCACTGGGACCTCAGAGTCTCTCTTGCACATCTTGG 143
 QY 126 CAATGACCTAAGCAATGCTGGGATAGCTGGAACCTGTTCTTCACGAGGACTTCAGAC 185
 DB 144 CAATGACCTAAGCAATGCTGGGATAGCTGGAACCTGTTCTTCACGAGGACTTCAGAC 203
 QY 186 AGTTGACAACTCTTAAAGCATCTCTTGAAGAACTCAAGTCCGACCTAGGAGTCTTCA 245
 DB 204 AGTTGACAACTCTTAAAGCATCTCTTGAAGAACTCAAGTCCGACCTAGGAGTCTTCA 263
 QY 246 GAAATCCAGTCTTGGCACTGGCCAGCAGAGAGGCCCGAGGAGCTGAGAAATTTGCTGA 305
 DB 264 GAAATCCAGTCTTGGCACTGGCCAGCAGAGAGGCCCGAGGAGCTGAGAAATTTGCTGA 323
 QY 306 CAATGTCATTTCTAAGTCTTCCAACTAACACGGACATTTTGGGTTCGAAATCAGCAA 365
 DB 324 CAATGTCATTTCTAAGTCTTCCAACTAACACGGACATTTTGGGTTCGAAATCAGCAA 383
 QY 366 CTCCTCATCTCTGATGTCAGGCTGAACGATCGATGATGCGCAAGGCTTAACTGAG 425
 DB 384 CTCCTCATCTCTGATGTCAGGCTGAACGATCGATGATGCGCAAGGCTTAACTGAG 443
 QY 426 CTCCTCATCTCTGATGTCAGGCTGAACGATCGATGATGCGCAAGGCTTAACTGAG 485
 DB 444 CTCCTCATCTCTGATGTCAGGCTGAACGATCGATGATGCGCAAGGCTTAACTGAG 503
 QY 486 GAAAGCTCTCTGACCTCTTCAACGCTGACATTTGAACTGATCCCAACACACCA 545
 DB 504 GAAAGCTCTCTGACCTCTTCAACGCTGACATTTGAACTGATCCCAACACACCA 563
 QY 546 GCCTGTTCCTGCTGGGAGATGGCCAGTGAACCCACACGATCTTCACTTTCCTGCT 605
 DB 564 GCCTGTTCCTGCTGGGAGATGGCCAGTGAACCCACACGATCTTCACTTTCCTGCT 623
 QY 606 GGACACACACGCAAAATCATCAAGTTCGTGAATAGCGTGTCAACACGCTGAAAG 665
 DB 624 GGACACACACGCAAAATCATCAAGTTCGTGAATAGCGTGTCAACACGCTGAAAG 683
 QY 666 CACTGTATCTCTCTGCTGCAAGGAGATATGTCCTGATGATCCGATCTTCACTC 725
 DB 684 CACTGTATCTCTCTGCTGCAAGGAGATATGTCCTGATGATCCGATCTTCACTC 743
 QY 726 CTTGATGTGATGTCATTCAGGAGTGGTGGTAAATCCCTCAGCACAAACCCAGCTGCA 785
 DB 744 CTTGATGTGATGTCATTCAGGAGTGGTGGTAAATCCCTCAGCACAAACCCAGCTGCA 803
 QY 786 AACCCCTCATTTGAAGAGGACCAATGAGGAGGACCACTGTGGTGCATGCTGTTTCC 845

Db 804 AACCTCATCTGAAGAGGACGAATGAGGAGGACCACTGTGGTCATGCTGATTGGTTCC 863
 QY 846 AGTGCTTGGCCACACCCCTTATAGCATCTCCCTCAGGAAGCTGCTGCCACCACTAAC 905
 Db 864 AGTGCTTGGCCACACCCCTTATAGCATCTCCCTCAGGAAGCTGCTGCCACCACTAAC 923
 QY 906 CAGCGTAAAGCCCTGAGTCTCCACGACGAGCTTCCACGATACCCCTTCTCCTCACAGT 965
 Db 924 CAGCGTAAAGCCCTGAGTCTCCACGACGAGCTTCCACGATACCCCTTCTCCTCACAGT 983
 QY 966 CAGAAGCAGGCTCTTACACATGTTCTGCTGCCCTGCGCAATAAAGGCCCATTTCTGCAA 1025
 Db 984 CAGAAGCAGGCTCTTACACATGTTCTGCTGCCCTGCGCAATAAAGGCCCATTTCTGCAA 1043
 QY 1026 AAA 1028
 Db 1044 AAA 1046

RESULT 2
 AF432917
 LOCUS
 DEFINITION Homo sapiens parotid secretory protein mRNA, complete cds.
 ACCESSION AF432917
 VERSION AF432917.1 GI:16755849
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Venkatesh, S.G., Geetha, C. and Gorr, S.-U.
 A member of the PSP/plunc family of BPI proteins is expressed in
 the human parotid gland
 Unpublished
 2 (bases 1 to 750)
 Venkatesh, S.G. and Gorr, S.-U.
 Direct Submission
 Submitted (12-OCT-2001) Molecular, Cellular and Craniofacial
 Biology, University of Louisville, 501 S. Preston Street,
 Louisville, KY 40292, USA

FEATURES
 source
 1. .750
 /location="Qualifiers"
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /sex="male"
 /tissue_type="parotid"
 1. .750
 /note="salivary protein; putative BPI protein; PSP;
 similar to BSP30 and plunc/lunx"
 /codon_start=1
 /product="parotid secretory protein"
 /protein_id="AAL28113.1"
 /db_xref="GI:16755850"
 /translation="MLQWLKLVLLCGVLTGTSESLDNLGNDLSNVVVKLEPVLHGL
 ETVDNTLKGLEKLVKLVGLQKSSAWQLAKQAEKLNINISKLLPTNTDIFGL
 KISNLSLDVKAEPIDDKGKLNLSFPVNTVAGPIGQIINKASLDLITAVTIEI
 DPQTHQPVAVLRCAADPTISLSLDKHSQIINKFVNSVINTLKSTVSSLLQKEICP
 LLRIFIHSLDWNVIQGVVDNPQHKTLQTLI"
 BASE COUNT 201 a 210 c 163 g 176 t
 ORIGIN

Query Match 72.6%; Score 746.8; DB 9; Length 750;
 Best Local Similarity 99.7%; Pred. No. 1.2e-192;
 Matches 748; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 49 ATGCTTTCAGCTTTGGAACCTTGTCTCTGTCGGCGTGTCTCACTGGGACTCAGAGTCT 108
 Db 1 ATGCTTTCAGCTTTGGAACCTTGTCTCTGTCGGCGTGTCTCACTGGGACTCAGAGTCT 60
 QY 109 CTTCTGACAATCTTGGCAATGACCTAAGCAATGCTGCTGGATAGCTGGAACTGTTCTT 168


```

QY 824 TGCTGATGCTGATTTGGTTCCAGTGGTTGCCACACCCCTTATAGCATCTCCCTCCAG 883
Db 787 GCCTGCTGACTGGTTCAGGACGTCACTGAATCTCTGTACCGTCT---TCTCTGGGACG 843
QY 884 GAAGCTGCTGCACACACCTAACACGCTGAAGCGCTGAGTCCACACGAGGACCTTCCC 943
Db 844 GTGCGTCTGCGCCACCATCCCGCAGGAGTGACAACCTGAGCCAGTCAAGAGACACTCTCAC 903
QY 944 AGATACCCCTTCTCTCTACAGCTCAGAACGAGGCTCTACACATGTTGTCTGCCCCCTGG 1003
Db 904 A-----CACTGCTCTACAGTTCAGGA---CACCTATGCTGTGTACACCTCCACCCCCAG 955
QY 1004 CAATAAGGCCCATTTCCCAAAA 1028
Db 956 CAATAAGAGCCCTTTTTCAGCAAAA 980

RESULT 4
LOCUS BTU79414 1024 bp mRNA linear MAM 09-DEC-1996
DEFINITION Bos taurus common salivary protein BSP30 mRNA, form b, complete
ACCESSION U79414
VERSION U79414.1 GI:1710368
KEYWORDS cow.
SOURCE Bos taurus
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 1024)
AUTHORS Haigh,B.J., Wilkins,R.J. and Wheeler,T.T.
TITLE The cloning and sequencing of two cDNAs coding for alternate forms
of BSP30, a bovine member of the Parotid Secretory Protein family
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1024)
AUTHORS Haigh,B.J., Wilkins,R.J. and Wheeler,T.T.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1996) Dairy Science Group, AgResearch, Private
Bad, Hamilton, New Zealand
FEATURES
source
1..1024
/organism="Bos taurus"
/db_xref="taxon:9913"
/tissue_type="parotid gland"
35..757
/notes="common salivary protein, form b"
/codon_start=1
/product="BSP30"
/protein_id="AAB38283.1"
/db_xref="GI:1710369"
/translation="MVQLKLVLLCGLLAGTSASIPDIRGNDVLRKLSGLRGLDTF
DSVLEIFQNLKTELESRCSDVEYEQETENFLEQLSRIFQVVSRLTGVRIRNVQVP
DITFEATSESNANLPIADVTSLPFLGEIVDLNVDLQTTVSIETDEDPQVY
GECTNNPESLSLVRHSRFLGNVDVIGNLARRVSSWEGELCPFRFRELLESIDA
ECVEKLIGESQDTTQPEGRS"
BASE COUNT 289 a 251 c 254 g 230 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
```


REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 974)

Strausberg, R.
Direct Submission
Submitted (05-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT

Contact: MGC help desk
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalob@bcm.tmc.edu
Villalob, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 16 Row: g Column: 1.

FEATURES

source

1..974
/organism="Mus musculus"
/db_xref="LocusID:19194"
/db_xref="taxon:10090"
/clone="MGC:11605 IMAGE:2616413"
/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma, 5 month old virgin mouse."
/clone_lib="NCI-CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
45..752
/codon_start=1
/product="parotid secretory protein"
/protein_id="AAH10288.1"
/db_xref="GI:16307481"
/translation="MFQGLSLVLCGLIGNSESLGELGSANNLKLINPPSEAVPQ
NLNDVLELQATSWPLAKNSILETLNTADLNKFTSLNGLLLKINNLKVLDFQAK
LSSNGGIDLVPLAGEASLVLPFGTIVGTVSLDLSLKTNAQGLPEVTIGK
CSSNPKISISLLGRRLPINSILDGVSLLTSLTVLQNFCLPQLLYVLINPSV
LQGLLSNLAQVOLAL"

CDS

260 a 258 c 181 g 275 t
Query Match 19.3%; Score 198; DB 10; Length 974;
Best Local Similarity 55.7%; Pred. No. 5.4e-43;
Matches 511; Conservative 0; Mismatches 360; Indels 47; Gaps 5;
Qy 40 AGACAAAGATGCTTCAGCTTTGGAAACTGTCTCTCTGCGGGCTCTCACTGGGACC 99
Db 36 AAACCAAGATGTTCCAACTTGGGAGCCCTGTGTCTGTGTGGCTCTCTCACTGGGACC 95
Qy 100 TCAGAGTCTCTTCTTGCAATCTTGCAATGACCTAAGCAATGTCTGTGATAGCTGCAA 159
Db 96 TCAGAGTCACTTCTTGCTGAACCTTGGCAGCGCTGTGAATAATTGA----- 141
Qy 160 CTTGTTCTTCACAGGAGCTTCAGACAGTTGACAACTCTTAAGGATCCTTCAGAAA 219
Db 142 -----AAATCTGAATCTCTCTCTGTAAGCTGTCTCTCAGAAC 179
Qy 220 CTGAAGGTGACCTAGGAGTGTCTCAGAAATCCAGATGCTTGSCAACTGGCCAAAGCAGAG 279
Db 180 CTGAATCTGATGTGGAGTGTCTTCAGCAAGCCACCAAGTTGGCCATTAGCCAAAGACAGC 239
Qy 280 GCCCAGAGAGCTCAGAAATGTGTACATGTCAATTTCTAAGCTGCTTCCCACTAACAGC 339
Db 240 ATTTCTAGAAACGTTGAACACACAGCGACCTTGGCAATTTAAAGAGCTTTACATCT---TTG 296

BASE COUNT
ORIGIN

RESULT 6
HSBA49G10 161593 bp DNA linear PRI 20-JUN-2001
LOCUS
DEFINITION
Human DNA sequence from clone Rpl1-49G10 on chromosome 20. Contains
a gene similar to bovine salivary protein BSP30, the LOC51297 gene
for LUNX protein, 5' end of a gene encoding a protein similar to
murine von Ebner minor salivary gland protein, a novel gene, a
pseudogene similar to STAT-induced STAT inhibitor-2, a pseudogene
similar to ribosomal protein L12, a putative novel transcript,
ESTs, STSS, GSSs and a CpG island, complete sequence.

ACCESSION AL121901.20 GI:8249854
VERSION
KEYWORDS HTG; BSP30; CpG island; LOC51297; LUNX; RPL12; STAT-inhibitor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 161593)
AUTHORS Tracey, A.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

COMMENT
On Jun 5, 2000 this sequence version replaced gi:8017404.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

IMPORTANT: This sequence is not the entire insert of clone RP11-49G10. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-49G10 is at 1 in this sequence. The true left end of clone RP5-1187J4 is at 161494 in this sequence. The true right end of clone RP4-733023 is at 27823 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

RP11-49G10 is from the library RPCI-11.1 constructed by the group of Peter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACes.6.

FEATURES

```

source
    Location/Qualifiers
        1..161593
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="20"
            /clone="RP11-49G10"
            /clone_lib="RPCI-11.1"
            complement(1..107)
                misc_feature
                    /note="match: GSS: Em:AQ607627"
                    6..706
                misc_feature
                    /note="match: GSS: Em:AQ051854"
                    137..548
                misc_feature
                    /note="match: GSS: Em:AQ550113"
                    530..603
                repeat_region
                    /note="37 copies 2 mer ca 66% conserved"
                    533..686
                repeat_region
                    /note="2 copies 77 mer 83% conserved"
                    613..678
                repeat_region
                    /note="33 copies 2 mer ca 75% conserved"
                    3030..3174
                repeat_region
                    /note="LIPB3 repeat: matches 5941..6084 of consensus"
                    5140..5183
                repeat_region
                    /note="22 copies 2 mer tt 75% conserved"
                    5184..5344
                repeat_region
                    /note="LIP2 repeat: matches 5996..6146 of consensus"
                    complement(6104..5463)
                misc_feature
                    /note="match: GSS: Em:AQ339300"
                    6120..6239
                repeat_region
                    /note="2 copies 60 mer 85% conserved"
                    complement(8476..9002)
                misc_feature
                    /note="match: GSS: Em:AQ537190"
                    complement(14766..15061)
                misc_feature
                    /note="match: STS: Em:HS922114S"
                    14772..14779
                misc_feature
                    /note="Random repeat. Forced join. Approximately 500 bases missing according to restriction digest."
                    16219..16367
                misc_feature
                    /note="match: GSS: Em:AL001671"
                    16225..16387
                misc_feature
                    /note="match: GSS: Em:AL013146"
                    18325..18826
                misc_feature
                    /note="match: GSS: Em:AQ458133"
                    18328..18828
                misc_feature
                    /note="match: GSS: Em:AQ532957"
                    18333..18770
            /note="match: GSS: Em:AQ765930"
            complement(20975..21245)
            /note="match: GSS: Em:A2102442"
            20999..21238
            /note="4 copies 60 mer 59% conserved"
            21031..21263
            /note="match: GSS: Em:A2055779 Em:A2084220"
            21144..21225
            /note="41 copies 2 mer ag 75% conserved"
            complement(21172..21428)
            /note="match: GSS: Em:AQ395327"
            complement(23473..23950)
            /note="match: GSS: Em:AQ727877"
            complement(23637..23948)
            /note="match: GSS: Em:AQ691469"
            23950..24355
            /note="match: GSS: Em:AQ152593"
            23955..24456
            /note="match: GSS: Em:AQ508832"
            26979..27422
            /note="match: GSS: Em:AQ636646"
            31882..32061
            /note="3 copies 60 mer 73% conserved"
            31921..32032
            /note="56 copies 2 mer ct 63% conserved"
            33693..33762
            /note="MER47 repeat: matches 2251..2322 of consensus"
            33733..33785
            /note="MER47 repeat: matches 1..55 of consensus"
            join(35535..35585,36507..36678,40308..40452,41455..41562,42783..42935,45523..45604,46980..47043,47896..47973,48600..48787)
            /gene="BA49G10.1"
            /note="match: cDNAs: Em:U79414 Em:U79413 Em:AX061621"
            match: ESTs: Em:A1654622"
            /product="BA49G10.1 (similar to bovine salivary protein BSP30)"
            evidence=not_experimental
            35535..48787
            /gene="BA49G10.1"
            join(36522..36678,40308..40452,41455..41562,42783..42935,45604,46980..47043,47896..47936)
            /gene="BA49G10.1"
            /note="match: proteins: Tr:P79125 Sw:P07743"
            /codon_start=1
            evidence=not_experimental
            /product="BA49G10.1 (similar to bovine salivary protein BSP30)"
            /protein_id="CAC03546.1"
            /db_xref="GI:9801234"
            /translation="MLQLKLVLCGVLTGTSESLLDNLGNDLSNVYDKLPVYLHGL
            ETVDTLKGLEKRLKVDLGVLOKSAWLAKAKAEKLLNNVSKLLPTNTDFGL
            KISNLLIDVKAEPIDDKGLNLSFPVTANTYVAGPIIGIINLKSLDLTLTET
            DPQTHQVAVLGEACASDPTISLSLLKHQSIIINKFNSVINTLKSTVSSLLQKEICP
            LIRPTHSLDVNVIIQVVVDNPOHQTQLTLI"
            38155..38751
            /note="LIP8 repeat: matches 4514..5125 of consensus"
            38750..39785
            /note="LIP8 repeat: matches 5128..6163 of consensus"
            48769..48774
            /gene="BA49G10.1"
            48787
            /gene="BA49G10.1"
            52540..53450
            /note="CpG island"
            evidence=not_experimental
            complement(53831..54131)
            /note="match: GSS: Em:AQ897745"
            53831..54141
            /note="match: STS: Em:G31731"
            53842..53969
            /note="match: GSS: Em:AQ429347"

```



```

Qy 829 CATGCTGATTGGTTCCAGTGGCTTCCGCCACCCCTTATAGCATCTCCCTCCAGAGC 888
Db 739 CCGTGGCTACTCTCAAGTGGTTTCATCTTGTGATGCAATTCCTCTAGAAAG- 797
Qy 889 TGTGTGCCACCACTAACCCAGCCTGAAGCCTGAGTCCCAACCAAGAGCCTTCCAGATA 948
Db 798 --TGCTACTATCATCACACATTTCTACCTGAG-ACCACCCAAAGGACCTCCCAAAAT 853
Qy 949 CCGCTTCTC 957
Db 854 CTCTTCTC 862

RESULT 8
RATPSP
LOCUS RATPSP 962 bp mRNA linear ROD 24-MAR-1998
DEFINITION Rattus norvegicus neonatal submandibular gland protein precursor
(psp) mRNA, complete cds.
ACCESSION M83209
VERSION M83209.1 GI:206456
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 962)
AUTHORS Mirels, L. and Ball, W.D.
TITLE Neonatal rat submandibular gland protein SMG-A and parotid
secretory protein are alternatively regulated members of a salivary
protein multigene family.
J. Biol. Chem. 267, 2679-2687 (1992)
JOURNAL
MEDLINE 9129360
REFERENCE 2 (bases 1 to 962)
AUTHORS Mirels, L., Miranda, A.J. and Ball, W.D.
TITLE Characterization of the rat salivary-gland B1-immunoreactive
proteins
Biochem. J. 330 (Pt 1), 437-444 (1998)
JOURNAL
MEDLINE 98129760
REFERENCE 3 (bases 1 to 962)
AUTHORS Mirels, L.
TITLE Direct Submission
Submitted (27-APR-1993) Molecular and Cell Biology, University of
California, 401 Barker Hall #3204, Berkeley, CA 94720-3204, USA
JOURNAL
FEATURES
source
1..962
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/cell_type="proacinar"
/tissue_type="submandibular gland"
/dev_stage="neonatal"
1..962
/gene="psp"
gene
1..962
/gene="psp"
CDS
47..754
/codon_start=1
/product="neonatal submandibular gland protein precursor"
/protein_id="AAC06334.1"
/db_xref="GI:206457"
/translation="MFQLGSLVLCGLLIGTSEILGDVANAVNLDILNPSSEAVAQ
NLNLDVGSLOQATTPSKADSLILEILNKVELGNSNGFTPLNGLLLRPNKFRVLQAG
LSNGKDIDLKPLVFETSFVPGTLDVAVSLLNLSVSVQNTAGTGLPGVTLGR
CSGNTDKISILIGRLRPLFPVNRILDGVSGLTGTAVSILLQNTLCPVLYLLSTWSGA
TQGLLSNLVTGLAVPL"
47..106
/sig_peptide
47..751
/gene="psp"
mat_peptide
107..751
/gene="psp"
/note="Secreted form"
/product="neonatal submandibular gland protein"
946..951
/polyA_signal
946..951
/gene="psp"

```

```

BASE COUNT 231 a 257 c 199 g 275 t
ORIGIN
Query Match 16.8%; Score 173.2; DB 10; Length 962;
Best Local Similarity 53.1%; Pred. No. 3.1e-36;
Matches 490; Conservative 0; Mismatches 388; Indels 44; Gaps 4:
Qy 38 CAAGACAAAAGATGCTTCAGCTTTGGAAACTTGTCTCTCTGCGCGTGTCTCACTGGGA 97
Db 36 CCAACCAAAAGATGTTCCAACTTGGAGCCTTGTGTCTGTGTGGCCTGTCTATTGGGA 95
Qy 98 CTTGAGAGTCTCTTCTTGCAATCTTGGCAATGACCAATGACCAATGCTGTGATAAGCTGG 157
Db 96 CTTGAGATCACCTCTGTGTGACGTTGCCAATGCTGTGATAATTTG----- 142
Qy 158 AACCTGTTCTTACGAGGAGCTTTCAGACAGTGTGACAAATCTCTTAAAGCATCTTTGAGA 217
Db 143 -----GACATTTCTAAATTTCTCCCTCTGAAGCTGTGCTCAGA 179
Qy 218 AACTGAAGTCGACCTAGGAGTCTTTCAGAAATCCAGTCTTGGCAACTGGCCAAAGCAGA 277
Db 180 ATCTGATCTGGATGTGGGTGCTACTTCAGACGCCACAACTTGGCCATCAGCCAAAGACA 239
Qy 278 AGCCCAAGAGCTGAGAAATGCTGACAAATGCTATTTCTTAAGTGTCTTCAACTAACA 337
Db 240 GCATTTCTTGAACATTTGAACAAAGTGGAGCTTGGCAATTTCAAAATGGCTTTTACACCT 296
Qy 338 CGGACATTTTGGTGTGAAATCAGCAACTCCCTCATCTCGATGTCAAAGCTGAACCGA 397
Db 297 TGAATGCTTTATTTGTCGAGTCAATAAATTTAGAGTCTCTGGATCTTCAAGCTGGCTGT 356
Qy 398 TCGATGATGGCAAGGCTTAACTGAGCTTCCCTGTACCGCGGAATGTCACTGTGGCGG 457
Db 357 CTTCCAATGCAAGGACATCGACTGAAAGTGCCTGTGTTTGAATTTCTTTTCTC 416
Qy 458 GGCCCATCTTTGGCAGATTATCAACTGAAAGCCTCTTGGACCTCTCGACGAGTCA 517
Db 417 TGCCTGCTATTGGCCTACGCTCGACGTGCTTTCCTTGGACCTCTTAACTCACTCA 476
Qy 518 CAATTGAACTTGATCCCGACACACAGCCTGTTCCTGCTGGGAGATGCGGCAGTG 577
Db 477 GTGTTCAAACCAATGCCAGACTGGCTCCTGGGTGACCTTAGGAAATGCTCCGTA 536
Qy 578 ACCCAACGAGCTCTACTTTCTCTGTCGACAAACACAGCCAAATCATCAACAGTTCG 637
Db 537 ATACAGACAAATCTCCATTTCTTGTGGCAGACGACTTCCCTTCGCAACCGTATTC 596
Qy 638 TGAATAGCGTATCAACACGCTGAAAAGCACTGTATCTCTCCCTGCTGCAGAGGAGATAT 697
Db 597 TGGATGCTATCTGGCTCCTCAGAGTGCAGTCAATCTCTGCTGCAAAACATCTAT 656
Qy 698 GTCCACTGATCCCGCATCTTCTACCTCCCTGGATGTGAATGTCTATTCAGCAGGTCTCG 757
Db 657 GCCCAGTGTCCAGTACTCTCTCAGCACAATGAGTGAAGTGTCTATTCAGGGTCTCCTCT 716
Qy 758 ATAATCTCT-AGCACAAAACCCAGCTGCAAAACCTCATTTTGAAGAGGAGGAATGAGGAG 816
Db 717 CTAATGATTAACAGGACAGCTAGCAGTCCCTCTCTGAAGAGGAGGAAGAGGAGAA 776
Qy 817 ACCACTGTGGTGTGATGTGTTGTTCCCGAGTGGCTTGGCCACCCCTTATAGCATCTC 876
Db 777 GCCATGACAAACCCCTGTGTTAGTCTCCACTGGCTTTATTTTACTTCGCGGTGGCAATTC 836
Qy 877 CCTCCAGGAAGCTCTGCCACCACTAACCCAGCGTGAAGCCCTGAGTCCCAACCAAGGA 936
Db 837 TTTCTGGGAAG----TTCTACTATATCCACAAAGTCTACCTGATACCAACGGAAGA 892
Qy 937 CTTTCCAGATACCCCTTCTCC 958
Db 893 CTTTCCAGATATCCCTTCTTTC 914

```

1
RESULT
LOCUS MPMSPG 13962 bp DNA linear ROD 29-OCT-1999
DEFINITION M.musculus Psp gene for parotid secretory protein.
ACCESSION X68699
VERSION X68699.1 GI:53809
KEYWORDS parotid secretory protein.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 13962)
Svendsen,P., Laursen,J., Krogh-Pedersen,H. and Hjorth,J.P.
Novel salivary gland specific binding elements located in the Psp
proximal enhancer core
Nucleic acids research. 26 (11). 2761-2770 (1998)
JOURNAL 98256451
MEDLINE 9592166
REFERENCE 2 (bases 1 to 13962)
AUTHORS Mikkelsen,T.R.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1992) T.R. Mikkelsen, Department of Molecular
Biology, University of Aarhus, Cf Mollers Alle 130, 8000 Aarhus,
DENMARK
FEATURES
source
Location/Qualifiers
1..13962 "Mus musculus"
/organism="Mus musculus"
/strain="C3H/AS"
/db_xref="taxon:10090"
/chromosome="2"
/germline
/clone_lib="Lambda-PHAGE (Lambda L47.1)"
/clone="Lambda Yp1, Lambda Yp3, Lambda Yp7"
/map="Estimate: 69 cm from centromere"
/note="Allele: b"
1..1578
evidence=experimental
4632..4678
/gene="psp"
/note="exon a"
/number=1
evidence=experimental
/product="parotid secretory protein"
4632..12433
/gene="psp"
5480..5608
/gene="psp"
/note="exon b"
/number=2
/product="parotid secretory protein"
6133..6274
/gene="psp"
/note="exon c"
/number=3
/product="parotid secretory protein"
6835..6942
/gene="psp"
/note="exon d"
/number=4
/product="parotid secretory protein"
7716..7868
/gene="psp"
/note="exon e"
/number=5
/product="parotid secretory protein"
10043..10124
/gene="psp"
/note="exon f"
/number=6
/product="parotid secretory protein"
10779..10842
/gene="psp"
/note="exon g"
/number=7
/product="parotid secretory protein"
11287..11361
/gene="psp"
/note="exon h"
/number=8
/product="parotid secretory protein"
12262..12433
/gene="psp"
/note="exon i"
/number=9
/product="parotid secretory protein"
3 others
BASE COUNT 3723 a 3075 c 3231 g 3930 t
ORIGIN
Query Match 6.9% Score 70.6; DB 10; Length 13962;
Best Local Similarity 60.8%; Pred. No. 3.7e-08;
Matches 115; Conservative 0; Mismatches 74; Indels 0; Gaps 0
QY 439 GCGAATGCTACTGTGGCGGGCCCATCATTTGGCCAGATTTATCAACCTGAAAGCCTCTTG 498
Db 7696 GTGACTCTGTTTCTGCGAGCGCTTCATTGGCAAGACGTCGACATTTCTGTTTCTTG 7755
QY 499 GACCTCTGACCGGAGTCACAAATTGAACCTATCCCGACAGACACACAGCCTGTTGCCGTC 558
Db 7756 GACCTCAATAATTCACCTCTCCATTAACCAATGCCAGATGCCCTTCTCTGAGTGACC 7815
QY 559 CTGGGAGATGGCGAGTCGACCCAGCATCTCATTCTTCTTGCTGGACAAACACAGC 618
Db 7816 ATAGCAAAATGCTCCAGTAATACAGATAAAATCTCCATTCTTCTTGGGAAGTAAGGC 7875
QY 619 CAAATCATC 627
Db 7876 TTATTCATC 7884
RESULT 10
RATSMGB
LOCUS Rattus norvegicus neonatal submandibular gland proacinar cell
DEFINITION protein precursor (SMGB1/SMGB2) mRNA, complete cds.
ACCESSION M83210
VERSION M83210.1 GI:206989
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 827)
Mirels,L. and Ball,W.D.
Neonatal rat submandibular gland protein SMG-A and parotid
secretory protein are alternatively regulated members of a salivary
protein multigene family
J. Biol. Chem. 267, 2679-2687 (1992)
JOURNAL 92129360
MEDLINE 9129760
REFERENCE 2 (bases 1 to 827)
AUTHORS Mirels,L., Miranda,A.J. and Ball,W.D.
TITLE Characterization of the rat salivary-gland Bi-immunoreactive
proteins
Biochem. J. 330 (Pt 1), 437-444 (1998)
JOURNAL 98129760
MEDLINE 9129760
REFERENCE 3 (bases 1 to 827)
AUTHORS Mirels,L.
TITLE Direct Submission
JOURNAL Submitted (27-APR-1993) Molecular and Cell Biology, University of
California, 410 Barker Hall #3204, Berkeley, CA 94720-3204, USA
FEATURES
source
1..827
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"

```

/strain="C3H.Amy-YBK"
/db_xref="taxon:10090"
/map="2"
/tissue_type="spleen"
<1..55
/gene="Psp"
/number=1
56..184
/partial
/gene="Psp"
/note="first expressed exon"
/number=2
64..123
/gene="Psp"
185..>210

```

RESULT	12
AE001965/c	
Locus	12334 bp DNA linear BCT 22-NOV-1999
DEFINITION	Deinococcus radiodurans R1 section 102 of 229 of the complete chromosome 1.
ACCESSION	AE001965 AE000513
VERSION	AE001965.1 GI:6458892
KEYWORDS	.
SOURCE	Deinococcus radiodurans.
ORGANISM	Deinococcus radiodurans Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcaceae; Deinococcus.
REFERENCE	1 (bases 1 to 12334)
AUTHORS	White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,

TITLE	JOURNAL	REFERENCE	AUTHORS
Genome sequence of the radioresistant bacterium <i>Deinococcus radiodurans</i> R1	Science 286 (5444), 1571-1577 (1999)	2 (bases 1 to 12334)	White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.

Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zaleski,C.,
Makarov,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
Smith,H.O., Venter,J.C. and Fraser,C.M.
Direct Submission
Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Location/Qualifiers
1. 12334
/organism="beinococcus radiodurans"
/strain="ri"
/db_xref="taxon:1299"
/chromosome="1"
complement(108. 1454)
/gene="DR1153"
complement(108. 1454)
/gene="DR1153"
/feature="DR1153"
/feature="similar to GB:U00096 SP:P77671 PID:1773192
PID:1786722 percent identity: 64.13; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="allantoinase, putative"
/protein_id="AAFI0727.1"
/db_xref="GI:6458893"
/translation="MSLDLLRLRGAVLVTPGERRADLGIVGGIAGIELTDEIATPAQOT
LDVGLHVPGLVDHHLNPEKTHWEGFETQALAGGATSFLOMPLNSPPVLT
REPERDKARGLDLDFGLWGGLTFLNLDQDLDAECGVIGLKAFMSHSLDEPFA
ADDTLYEGTRAKRHLVATHAESNEFTRRTETARAGKSCVRODYLESPPVTEL
EAVQALLFAQDTGAALLVHVSSGAVALAYEGKOKGIDVTIETCPHYLHFTGEDE
RVGAALCAPDPAQVQBELWRELLAGHIDTQSDSPAPDMKTSDEFFSLWGGIS
GAQSTLNMPLDGYAQRGLPLEIIAALLNPAQRGLPKQGLAVGADADALVALG
EKFTLDLYRWQONPYRQGSFQGRVHATVLRQPYQNGEFTGTPRGLLRPSL"
complement(1451. 2701)
/gene="DR1154"
complement(1451. 2701)
/gene="DR1154"
/feature="similar to SP:P37113 GB:S67784 GB:X74289 PID:460895
SP:053389 percent identity: 55.20; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="N-carbamyl-L-amino acid amidohydrolase"
/protein_id="AAFI0728.1"
/db_xref="GI:6458894"
/translation="MNETQAOLTROVLDTCALACHTEVPGEITETLYLCPTRQVHE
ELRWARLGMTTYDAGVNLRSLSKAPGARTLYIGSHLDIVPAGRIDGIGVVF
GYLVAERLELPFHLVGLFSEERGVYGVFGISRALVGTADELLTVSDRENQTV
RDAIVGGLNDELPGAGEDRPGLYLETHIEQGPVLQDQGAAGVSVIAVGSRLTL
HFTGRASHGTTPHLKRDAALAAAREIVGAEDLANRTPGLVATVGMIEAKPGAGNVI
AGEVSCSIDIRHADDVRAQSLQELLALAEARAAQVACSVTPRMAEAAVPMAPSLI
ALLHQAEEQGLTHPELVSGAGHDAQIMAKMPPAAMFLRSPNALSHHPDMAEPQDV
AAGLVGTGTFLELLAAQEQVQA"
complement(2698. 4089)
/gene="DR1155"
complement(2698. 4089)
/gene="DR1155"
/feature="similar to SP:P08997 GB:X12431 PID:396349 PID:40887
GB:U00096 percent identity: 49.74; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="malate synthase, putative"
/protein_id="AAFI0729.1"
/db_xref="GI:6458895"
/translation="MNPPLSLSPAALALARQLHRLRWRAGLPGEPARANSVLTA
ARAAIPDRLGRRAELIVASDLMLRRVADSDADAVVDFDDTSPTRDNVQASD
ALAWAESDEPLARALYALPHLDFDGPAATAALCDLSAVLAARPERPHLYPKL
ETPAEQFWHDLALLAEHLSLAPNTVKTCLQIETPSGLNINEAVEPELAPRYGLNA
GRWDIVSLTKHLGTRDTEMPFRSLTMDVDAMRAYEFLVRCRAHGAEEIGGTGAS
FADPQNSQALGAVLADKQRAAQGTAAWAGLPELLEPVREGRFADPAQSNKQTPA
DYLHRLTDLPGPLGRADREALIETALGVFAAWYREGVLRAGRIEDTATAELARA
QLQWQVRYRALEDGERLTPERYQRERRAVLGDDPEASRLDLHLVAPACAAVFPLEA

gene
CDS
gene
CDS
gene
CDS
gene
CDS

QRLHLSDSSEKPE"
complement(4153. 4956)
/gene="DR1156"
complement(4153. 4956)
/gene="DR1156"
/feature="similar to SP:P37728 GB:X62072 PID:48982 percent
identity: 58.17; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="transcriptional regulator"
/protein_id="AAFI0730.1"
/db_xref="GI:6458896"
/translation="MTEAPKLKSGRARSGEAGSVRTLGRGLSVLSALAELEATLTQV
AKAGLSASTTYRLLETLRQCGYVWEERSGLFSVLRAYQVGAFAEFERNLSLSAAGA
VMOALVEEVNETANLAVLNRQYVYHQVEASOLMRMFTHIGAAAPLHCSCGVCKVLT
WLPAAEVRKVGDCGPPAYTPHSITTLAALTRELTNYRGODYALDDEEREELGYRCLAT
PVRAAGEVVASLSVSAPTSRFPKKNIPDMLGRVQAASEQISARLGRG"
complement(5161. 6392)
/gene="DR1157"
/feature="This region contains an authentic frame shift and
is not the result of a sequencing artifact; similar to
PID:805293 percent identity: 62.08; identified by sequence
similarity; putative; hydroxypruvate reductase, putative,
authentic frameshift"
6657. 7223
/gene="DR1158"
6657. 7223
/gene="DR1158"
/feature="similar to GB:AL009126 percent identity: 58.14;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAFI0731.1"
/db_xref="GI:6458897"
/translation="MTRPLTLEOLNALSDDAFTEHPAGVLEHSPHVARAAAGRPFA
DVEEAAAFARAVADPEGAQVOLIHAHPDLACKAALAGELTAESAEOPTSAGLDRLS
PEEAAEQRLNAAHYHERFGLPYVVCYRENTKDTIFEGARRRLTHQEEQAAALHEIG
PIARLRLDLVQEGRQERKMDGRLIRP"
complement(7360. 7905)
/gene="DR1159"
complement(7360. 7905)
/gene="DR1159"
/feature="similar to GB:X74409 SP:P42195 PID:396458 percent
identity: 56.72; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="transcriptional regulator, MarR family"
/protein_id="AAFI0732.1"
/db_xref="GI:6458898"
/translation="MSARMNDTAAALLERISDWARLNHGQGPDSGLTSPSAGPMLTL
LLERLHAALGREIERTYAASGLNAGWDLTLTYSAPPEGLRPTLSALAAISGTS
TSNRIVRLLEKGLIERREDERDRRSIRLTPQGRALVTHLLFAHLATTQRLAPLSA
QEQRTEELAGRLAGLEQGV"
8007. 8903
/gene="DR1160"
8007. 8903
/gene="DR1160"
/feature="similar to SP:P34798 PID:432629 percent identity:
57.45; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="uricase"
/protein_id="AAFI0733.1"
/db_xref="GI:6458899"
/translation="MMTGQTPQPKVYKVRLENNGYKAEVQMLKIKRGTTPHREURE
AKVRVYMGDFAAHSECDNTLVATDVRNTVYVYLAKEGFESSIEFGKELLTHFVK
VPRVTGCGFAFTEHLVQVTPAQPGQHDHAFVQPKRTARVETQDGRRTVTSGI
EELVLTDESGWENYLLDERFTLTPETHDRVNAVFTAKVAVESCYDYDAWERYV
RQIOHTFDHSPSLQRTLVLMGEAVLSVCPSEIRWIFQMPKNHLYVNLGRGLENN
NEILHVPPEPYGLMEANVERAE"

bb 460 AGCTCAAGTGAATACGCCCTGGCTGCTGCAAGTCTGTTGAGGCTGGCTGTGAAGCTGG 519
yy 503 TCCTGACCGCAGTCAAAATGAAACTGATCCCCACACACACACAGCCTGTTGCCGTCTCTGG 562
bb 520 ACATCACAGCAGAAATCTTAGCTGTGAGAGATAAGCAGGAGAGGATCCACCTGGTCTCTTG 579
yy 563 GAGAAATGCCCGCAGTGACCCCAACACAGCATCTCTACTTTTCTTGCTGGACAAACACAGCCAAA 622
bb 580 GTGACTGGACCCCATTCGCCCTGGAAGCCTGCAAAATTTCTCTGCTTGATGGACTTGGCCCCC 639
yy 623 TCATCA 628
bb 640 TCCCA 645

Search completed: August 6, 2002, 17:26:32
Job time: 5124 sec

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2002, 14:03:33 ; Search time 1685.42 Seconds
(without alignments)
8232.292 Million cell updates/sec

Title: US-10-020-139-1

Perfect score: 1028
Sequence: 1 CACGACATTTCATGACATC.....AAGGCCATTTCTGCAAAA 1028

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

EST:
1: em_estba:
2: em_esthum:
3: em_estin:
4: em_estmu:
5: em_estnu:
6: em_estpi:
7: em_estro:
8: em_hic:
9: gb_est1:
10: gb_est2:
11: gb_hic:
12: gb_gss:
13: em_gss_hum:
14: em_gss_inv:
15: em_gss_pln:
16: em_gss_vrt:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	341.4	33.2	384	9	A1654622
2	218	21.2	253	9	A1218468
3	200.8	19.5	891	10	BG869528
4	199.8	19.4	896	10	BG870334
5	198.4	19.3	844	10	BG872145
6	198	19.3	972	11	AK009032
7	198	19.3	972	11	AK010181
8	196.8	19.1	954	10	BG872451
9	196.4	19.1	971	11	AK009654
10	195.8	19.0	892	10	BG865328
11	195.2	19.0	870	10	BG865874
12	193.6	18.8	968	10	BG870337
13	192	18.7	910	10	BG872403
14	191.2	18.6	1039	10	BG865758
15	190.8	18.6	889	10	BG868020
16	190.8	18.6	982	10	BG869767
17	189.6	18.4	796	10	BI688385

18	189.6	18.4	890	10	BG870690
19	189.2	18.4	1050	10	BG870257
20	189	18.4	867	10	BG872038
21	189	18.4	934	10	BG871619
22	188.8	18.4	920	10	BG871297
23	188.4	18.3	857	10	BG867154
24	188.2	18.3	740	10	BG872659
25	188	18.3	797	10	BG867408
26	188	18.3	933	10	BG872825
27	187.8	18.3	878	9	AA792682
28	187.8	18.3	943	10	BG869519
29	187.6	18.2	834	10	BG869465
30	187.6	18.2	967	10	BG867259
31	187.6	18.2	992	10	BG870228
32	187.4	18.2	948	10	BG869588
33	187.4	18.2	955	10	BG871442
34	186.6	18.2	781	10	BF535510
35	186.6	18.2	1029	10	BG855861
36	186.2	18.1	778	10	BG865641
37	186.2	18.1	824	10	BG866924
38	186.2	18.1	833	10	BG869454
39	186.2	18.1	843	10	BG872130
40	186	18.1	881	10	BG871864
41	185.6	18.1	928	10	BG872117
42	185.6	18.1	955	10	BG870254
43	185.4	18.0	907	10	BG871444
44	185	18.0	860	10	BG870889
45	184.8	18.0	707	10	BG866908

ALIGNMENTS

RESULT 1
AI654622/C

LOCUS

AI654622 wb48f11.x1 NCI_CGAP_GC6 Homo sapiens CDNA clone IMAGE:2308941 3',
384 bp mRNA linear EST 17-DEC-1999

DEFINITION

AI654622 mRNA sequence.

ACCESSION

AI654622

VERSION

AI654622.1 GI:4738601

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 384)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone Distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bioinformatics.nih.gov/bbrp/image/image.html

Insert Length: 472 Std Error: 0.00

Seq primer: -400P from Gibco.

Location/Qualifiers

1. 384

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2308941"

/clone_lib="NCI-CGAP-GC6"

/tissue_type="pooled germ cell tumors"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA

from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaudo.

BASE COUNT 81 a 72 c 119 g 112 t
ORIGIN

Query Match 33.2%; Score 341.4; DB 9; Length 384;
Best Local Similarity 96.9%; Pred. No. 1.6e-80;
Matches 348; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 670 GTATCCCTCCCTGTCGAGAGGAGATATGCCAGTATCCCGATCTTTCATCCATCCCTG 729
Db 384 GTATCCCTCCCTGTCGAGAGGAGATATGCCAGTATCCCGATCTTTCATCCATCCCTG 325
Qy 730 GATGTGATGTCATTCAGCAGGTGTCGATATCTCTCAGCACAACCCAGCTGCAACCC 789
Db 324 GATGTGATGTCATTCAGCAGGTGTCGATATCTCTCAGCACAACCCAGCTGCAACCC 265
Qy 790 CTCATTTGAAGAGGAGATGAGGAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 849
Db 264 CTCATTTGAAGAGGAGATGAGGAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 205
Qy 850 GCTTGCCCAACCCCTTATAGCATCTCCCTCCAGGAGTGTGCTGCTGCTGCTGCTGCTG 909
Db 204 GCTTGCCCAACCCCTTATAGCATCTCCCTCCAGGAGTGTGCTGCTGCTGCTGCTGCTG 145
Qy 910 GTGAAGCCCTGAGTCCACACAGAGGACCTTCCAGATACCCCTTCTCTCAGATCAGA 969
Db 144 GTGAAGCCCTGAGTCCACACAGAGGACCTTCCAGATACCCCTTCTCTCAGATCAGA 85
Qy 970 ACAGCAGCTCTACATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1028
Db 84 ACAGCAGCTCTACATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26

RESULT 2
A1218468/C
LOCUS qh27e04.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:1845918 3', mRNA sequence.

ACCESSION A1218468
VERSION A1218468
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 472 Std Error: 0.00
Seq primer: -400p from G1bco
High quality sequence stop: 207.

FEATURES
source
1..253
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1845918"
/lab_host="Soares_NFL_T_GBC_S1"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI;

Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBL19W, testis NHT, and B-cell NCI_CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaudo.

BASE COUNT 48 a 48 c 77 g 80 t
ORIGIN

Query Match 21.2%; Score 218; DB 9; Length 253;
Best Local Similarity 97.8%; Pred. No. 1.2e-47;
Matches 221; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 803 GAGCAATGAGGAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 862
Db 253 GAGCAATGAGGAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 194
Qy 863 CCTATAGCATCTCCCTCCAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 922
Db 193 CCTATAGCATCTCCCTCCAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 134
Qy 923 TCCACACAGAGGACCTTCCAGATACCCCTTCTCTCAGATCAGACAGCAGCTCTA 982
Db 133 TCCACACAGAGGACCTTCCAGATACCCCTTCTCTCAGATCAGACAGCAGCTCTA 74
Qy 983 CACATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1028
Db 73 CACATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 28

RESULT 3
BG869528
LOCUS 602789545F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4921042 5',
DEFINITION mRNA sequence.

ACCESSION BG869528
VERSION BG869528.1 GI:14220068
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS J (bases 1 to 891)
TITLE NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM10838 row: i column: 11
High quality sequence stop: 834.

FEATURES
source
1..891
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4921042"
/lab_host="NCI_CGAP_SG2"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Constructed by Life

D _b	423	CTTTTCAATGGCAGACGGTCGACATTCTGTTCCTTGGAACCTCATAAATTCACCTCC	482
Q _y	520	ATTGAAACTGATCCCCAGACACACAGCCCTGTGCCGTCTGGGAGAATGCCGCAGTAGC	579
D _b	483	ATTAAAACAATUCCCAGACTGGCCCTCTCTGAGGTGACATAGGCAAAATGCTCCAGATAA	542
Q _y	580	CCAACCAGCATCTCACTTCTCTGCTGGACAACACGCCAAATCATCAACAAGTTTCGNG	639
D _b	543	ACRGATAAATCTUCATTTCTTGTTGGAGACGANTTACCCTATCATCACAGATTCTTG	602
Q _y	640	AATAGCGTGATCAACACGCTGAAAGACACTGTATCTCCCTGCTGCAGAGGAGATATGT	699
D _b	603	GATGGTGCTCTACCCCTCCCTTACAAGTACATTGTCAACGGTCCCTGC AAAACTTCCTATGT	662
Q _y	700	CCACTGATCCGCATCTTCATCCCACTCCCTGGATGTGAATGTCAATTCAGCAGGTCGTCGAT	759
D _b	663	CCATTGGCTTCAGTAGTCTCTACGACACTGAATCCAAUGTTCCTTCAGGSTCTCCCTCTCT	722
Q _y	760	AATCTCTAGCACAAAAACCCAGCTGCAAAACCCCTCATTTTGAAGAGGACGAATGAGGAGGACC	819
D _b	723	AATCTATTGG--UCAGAGACAGGTACAACCTGGCCCTCTGAAGAGGAAGAACAGAGGGATG	779
Q _y	820	ACTGTGGTGCATGCTGATTGGTTCCCACTGGGCTGCCCCACCCCTTATAGCATCTCCCT	879
D _b	780	CACAACTCTCTCTGCTACTCTCCAGTGGTTTCATCTTACTTCTGATGSCATTTCCCT	839
Q _y	880	CCAGGAAGCTGCT	892
D _b	840	CTTAAGAAAGTCT	852

RESULT	5	
LOCUS	844 bp	linear
DEFINITION	603795983F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4924053 5', mRNA sequence.	
ACCESSION	BC872145	
VERSION	BC872145	
KEYWORDS	BC872145.1	GI:14222685
EST.		
SOURCE	house mouse	
ORGANISM	Mus musculus	
REFERENCE	1 (bases 1 to 844)	
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL	Unpublished (1999)	
COMMENT	Contact: Robert Strausberg, Ph.D.	

BASE COUNT
ORIGIN

212 a 214 c 172 g 246 t
/note="organ: salivary gland; Vector: pCMV-SPORT6; Site.1:
Not1; Site.2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
212 a 214 c 172 g 246 t

Query Match	19.3%	Score 198.4	DB 10	Length 844	
Best Local Similarity	55.5%	Prod. No. 3.9e-42			
Matches 471	Conservative	0	Mismatches 336	Indels 41	Gaps
Qy	40	AGACAAAGATGCTTCAGCTTGGAAACTTGTTCCTGTGGCGGTGCTCACTGGGACC	99		
Db	32	AAACCAAGATGTTCCAACTTGGAGCCCTGTGTCTTGTCTGGCCTGCTCATTTGGAAC	91		
Qy	100	TCAGAGTCTCTTCTTGACATCTTGGCATGACCTTAAGCAATGTGCTGGATAGACTGGAA	159		
Db	52	TCAGAGTCACTTCTGTGTGAACCTTGGCAGCGCTGTGCAATAATTGA	137		
Qy	160	CGTGTCTTCAGGAGGACTTGAACAGTGTGACATACCTCTTAAAGGCATCCTTGGAAA	219		
Db	138	-----AAATCTGAATCTCCCTCTGAAGCTGCCCTCAGAC	175		
Qy	220	CTGAAGTCGACCTAGGAGTGCTTTCAGAAATOCAGTGTGTGGCACTGGCCAGCAAG	279		
Db	176	CTGAATCTGGATGTGGAGTGTCTTCAGCAAGCCACAAGTTGGCCATTAGCAAGAAC	235		
Qy	280	GCCAGGAAGCTGAGAAATTCGTGAACAATGTCAATTTCTTAAGCTGCTTCCAACCTAACACG	339		
Db	236	ATTCTAGAAACGTTGAACACAGCGGACCTTGGCAATTTAAAGCTTTACATCT	292		
Qy	340	GACATTTTGGGTTGAAATCAGCAACTCCCTCATCTCTGGATGTCAAAGTGAACCGATC	399		
Db	293	AATGGTTTATTCTTGAAATCAATAATCTCAAAGTCCCTGGATTTTCAAGCTTAAGCTGTCT	352		
Qy	400	GATGATGGCAAGGCCCTTAACCTCAGCTTCCCTGTACCGGCAATGTCACTGTGGCGGG	459		
Db	353	TCCACGGCATATGCAATTTGATCTTGACGGTGCCCTTGGCTGGGGAAGCCCTCTTGTTCTG	412		
Qy	460	CCCATCATTTGCCAGATTATCAACCTGAAAGCCCTCCTTGGACCTCTGACCGCAGTCACA	519		
Db	413	CGTTTCATTTGGCAAGACGGTCGACATTTCTGTCTTCTTGGACCTCATAAATTCACCTCC	472		
Qy	520	ATTGAACTGATCCCGACAGACACACCGCTGTGGCGTCTCTGGGAGATGCCGACGTGAC	579		
Db	473	ATTAAACCAATGCCCAGACATGGCGCTTCTCTGAGTGACCATAGGCAAAATGCTCCAGTAAT	532		
Qy	580	CAACCCAGCATCTCACTTTCCCTGCTGGACAAACACAGCCAAATCATCAACAGTTCTGTG	639		
Db	533	ACAGATAAAAATCTCCATTTCCCTTGGGAAGACGATTACCCATCATCAACAGATTCTG	592		
Qy	640	AATAGCGTGATCAACACGCTGAAAGCACTGTATCCTCCCTGCTGCAGAGGAGATATCT	699		
Db	593	GATGGTGTCTCTACCCCTCTTACAGATGACATTTGTAACCGTCTGCAAAAGTTCCCTATGT	652		
Qy	700	CACTGATCCGCATCTTCATCCACTCCCTGGATGTGAAATGTCATTTACAGACGGTCTGAT	759		
Db	653	GCATTTGCTTTCAGTAGCTCTCAGCACACTGAATCCAAAGTGTCTTTCAGGGTCTCCTCTCT	712		
Qy	760	AATCCTCAGCAAAACCCAGCTGCAAAACCTCATTTGAGAGGACGAAATGAGGAGACC	819		
Db	713	AATCTACTGGGAGGA--CAGGTACAACCTTGGCCCTCTGAAGAGGAAGCAAGAGGATGC	770		
Qy	820	ACTGTGGTGCATGTGATTGGTTCCCAAGTGGCTTGCCCCACCCCTTATAGCATCTCCCT	879		
Db	771	CACAACTCTCTGCTGTACTCTCCAGTGGTGTTCATCTTACTTCTGATGGCATTCCTCT	830		
Qy	880	CCAGGAAG	887		
Db	831	CTAGAAAG	838		

RESULT 6
AK009032
LOCUS
DEFINITION

Db 236 ATTCTAGAAACCTTCAACACACAGCGACCTTGGCAATTTAAAAAGCTTTACATCT---TTG 292
 QY 340 GACATTTTGGGTTCAGAACTCAGCAACTCCCTCATCTCTGGATGTCAAGCTGACCCGATC 399
 Db 293 AATGGTTTATTGTTGAAATCAATATCTCAAGTCCCTGGATTTTCAAGCTAAGCTGTCT 352
 QY 400 GATGATGGCAAGGCTTAACTGAGCTTCCCTGTCAACCGGGAATGTCACCTGCGCCGG 459
 Db 353 TCCACGGCAATGATGATCTGACGGTGCCCTTGGCTGGGAAGCCTCTCTGGTCTG 412
 QY 460 CCATCATTTGGCCAGATATCAACCTGAAAGCCTCTTGGACCTCTTGGACCGGAGTCA 519
 Db 413 CTTTTCATTTGGCAAGCGGTCACATTTCTCTTCTTGGACCTCTATAAATTCACCTCC 472
 QY 520 ATTGAACCTGATCCCGACAGACACAGCCTCTTGGCCCTCTTGGGAGATGCGCCAGTGAC 579
 Db 473 ATTAAACCAATGCCAGACTGGCCCTCTCTGAGGTGACCATAGGCAATGCTCCAGTAAT 532
 QY 580 CCAACAGCATCTCACTTCTCTTCTGTCGACAAACACAGCCGAATCATCAACAAGTTCTGT 639
 Db 533 ACAGATAAATCTCCATTTCTTGTGGAAGAGCATTTACCCATCATCAACAGATTTCTG 592
 QY 640 AATAGCCTGATCAACACAGCTGAAAGAGCACTGTATCTCTCTCTGCTGCGAGAGGATATGT 699
 Db 593 GATGGTCTCTACCTCTTACAGATGATGTCACCGCTCTGCAAAACTTCTCTATGT 652
 QY 700 CCAGTATCCGATCTCATCTCCCTGATGTCATGTCATGTCATGTCATGTCATGTCATGTC 759
 Db 653 CCATGCTTCACTAGTCTCTCAGCACATGTAATCCAGTGTTCTTCAAGGGTCTCTCTCT 712
 QY 760 AATCTCAGCAACAAACCCAGCTGCAAAACCTCTATTTGAAGAGGACGAATGAGGAGACC 819
 Db 713 AATCTAATCTGGCAGGACAGATGATGTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 772
 QY 820 ACTGTGGTGCATGCTGATGTTGGTTTCCCATGCGCTTGC-CCCACCCCTTATAGCATCTCCC 878
 Db 773 CACAAATCTCTCTGCTGGGTACTCTCCAGTGGCTTTCTATCTTACTTCTGATGGCATTTCCC 832
 QY 879 TCCAGGAAGCTCTGCCACCA 899
 Db 833 TCTAGAAAGTCTTATCA 853

RESULT 11

BG865874 602788106F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4914193 5',
 LOCUS 870 bp mRNA linear EST 29-MAY-2001
 DEFINITION mRNA sequence.

ACCESSION BG865874
 VERSION BG865874.1 GI:34216414
 KEYWORDS EST.

SOURCE

house mouse.
 Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)

AUTHORS

Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov

TITLE

Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.

JOURNAL

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.

COMMENT

Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10820 row: 1 column: 02

FEATURES

High quality sequence stop: 858.
 Location/Qualifiers
 1..870

BASE COUNT 223 a 223 c 172 g 252 t
 ORIGIN
 Query Match 19.0%; Score 195.2; DB 10; Length 870;
 Best Local Similarity 55.3%; Pred. No. 2.9e-41;
 Matches 469; Conservative 0; Mismatches 338; Indels 41; Gaps 3;
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4914193"
 /clone_lib="NCI_CGAP_SG2"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
 NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
 dT. Average insert size 1.3 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

QY 40 ACACAAAGATGCTTCAGCTTTGGAAACTTGTCTCTCTGTGGGGGCTGCTCACTTGGGACC 99
 Db 25 AAACCAAGATGTTCCAACTTGGGAGCCTTGTGTCTTGTGGCCCTGCTCATTTGGGAAC 84
 QY 100 TCAGAGTCTCTTTCGACAATCTTGGCAATGACCTTAAGCAATGCTGCGTGAAGCTGGAA 159
 Db 85 TCAGAGTCACTTCTTGGTCAACTTGGCAGCGCTGTGAATAATTGA----- 130
 QY 160 CCUGTCTTCACGAGGACTTGAACAGCTTGAACAATCTCTTAAAGGCATCCTTGAGAAA 219
 Db 131 -----AAATCTGAATCT 168
 QY 220 CTGAAGTGCACCTAGGAGTGTCTCAGAAATCCAGTGTCTTGGCAACTGCCCAAGCACAAG 279
 Db 169 CTGAATCTGATGTGGAGTTGCTTCAGCAAGCCACAAGTTGGCCATTAGCCAAAGACAGC 228
 QY 280 GCCCAGGAAGCTGAGAAATGCTGAACAATGTCTTCTTAAGTGTCTTCCAACTTAACAGC 339
 Db 229 ATTCTAGAACCTTGAACACAGCGGACCTTGGCAATTTTAAAGAGCTTTTACATCT---TTG 285
 QY 340 GACATTTTGGGTGGAATATCAGCAACTCCCTCATCTCTGGATGTCAAAGCTGAACCGATC 399
 Db 286 ANTGGTTATTGTTGAAATCAATAATCTCAAAGTCTCTGGATTTCAGCTTAAGCTGTCT 345
 QY 400 GATGATGGCAAGGCTTAACTGAGCTTCCCTGTACCGCGGAATGTCTACTTGTGGCCGG 459
 Db 346 TCCACGGCAATGGCATGTCTGACGGTGGCCCTTGGCTGGGGAAGCCTCTCTTGGTCTG 405
 QY 460 CCATCATTTGGCCAGATATCAACCTGAAAGCCTCTTGGACCTCTCTCAAGCCAGTCA 519
 Db 406 CTTTTCATTTGGCAAGCGGTGACATTTCTGTTCTTGGACCTCATATAATTCACCTTCC 465
 QY 520 ATGAACTGATCCCGACAGACACAGGCTGTGTGCGGTCTCTGGGAGAAATGGCCAGTAC 579
 Db 466 ATTAAACCAATGCCAGACTGGCCCTCTCTGAGGTGACCATAGGCAAAATGCTCCAGTA 525
 QY 580 CCAACAGCATCTCACTTCTCTTCTGCTGGCAACACAGCCAAATCATCAAGTTCGTG 639
 Db 526 ACAGATAAATCTCCATTCTCTTGGTGGGAAGACGATTACCCATCATCAACAGTATTCTG 585
 QY 640 AATAGCTGATCAACAGCTGAAAGCACTGTATCTCTCTCTGCTGCAAGAGGAGATATGT 699
 Db 586 GATGGTGTCTCTACCTCTCTTACAAGTACATTGTCAACCGCTCTGCAAAACTTCTCTAT 645
 QY 700 CCATGATCCGATCTTCACTCACTCTCTGATGTGAATGTCATTCAGCAGGTCGTGAT 759
 Db 646 CCATGCTTCACTAGTCTCTCTCAGCACACTGAATCCAAAGTGTCTTTCAGGGGTCTCTCT 705
 QY 760 AATCTCTCAGCAACAAACCCAGCTGCAAAACCCCTCATTTGAAGAGGAGCAATGAGGAGAC 819
 Db 706 AATCTACTGGCAGGACAGGTACAACCTTGGCCCTCTCAAGAGGAGAGA--ACAGAGGATGCC 763
 QY 820 ACTGTGGTGCATGCTGATTTGTTCCAGTGGCTTGGCCCAACCCCTTATAGCATCTCCCT 879
 Db 764 ACAACTCTTCTCTGCTGGTACTCTCCAGTGGTTTCACTTACTTCTGATGGCATTTCCCT 823


```

source                                1..910
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/cdname="IMAGE:4924056"
/clone_lib="NCI_CGAP_SQ2"
/lab_host="DH10B (n1 phase-resistant)"
/notes="Organ: salivary gland; Vector: pCMV-SPORT6; Site1:
Not1; Site2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT      236 a   235 c   181 g   258 t
ORIGIN

Query Match      18.7%; Score 192; DB 10; Length 910;
Best Local Similarity 53.9%; Pred. No. 2.1e-40;
Matches 467; Conservative 0; Mismatches 360; Indels 39; Gaps 2;

QY  40 AGACAAAGATGCTTCAGCTTGGAAACTTGTCTCTCTGTGGCGCTGCTCACTGGGACC 99
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  29 AAACCAAGATGTTCCAACTTGGGAGCTTGTGGTCTTGTGTGGCTGCTCATTTGGGAAC 88

QY  100 TCAGAGTCTCTCTTTCAGACATCTTTGGCAATGACCTAAGCAATGCTGTGGATAAGCTGAA 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  89 TCAGAGTCACTTCTTGTGTAACCTTGGCAGCGCTGTGAATAATTGA----- 134

QY  160 CTTGTTCTTCAGAGGGGACTTGCAGACAGTTTGACAATACTCTTAAGGCATCTTGAGAAA 219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  135 -----AAATCTGAATCCTCCCTCTCAAGCTGTCCCTCAGAAC 172

QY  220 CTGAAGTTCGACTTAGGAGTGCTTTCAGAAATCCAGTGTTCGCAACTGCCCAAGCAGAAG 279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  173 CTGAATCTGGATGTGGAGTTCCTTCAGCAGCCACAAAGTTGGCCATTAGCCAAAGAACAGC 232

QY  280 GCCAGGAAGCTGCAAAATCTCTGAACAAATGTCTATTTCTAAGTGTCTCCAACTACACAG 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  233 ATTCTAGAACGCTTGAACACAGCGGACTTGGCAATTTAAAGAGCTTTACATCT---TTG 289

QY  340 GACATTTTGGGTTTGAANAATCAGCAATCCCTCATCTCTGGATGTCAAAGCTCAACCGATC 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  290 AATGGTTTATTGTGAAANAATCAATAATCTCAAAGTCTGCTGATTTTCAAGCTAAGCTGTCT 349

QY  400 GATGATGGCAAGGCTTTAACTCTAGCTTCCCTCTCACCGCGAATCTCACTGTGCCCGGG 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  350 TTCACGGCAATGGCAATGTCTGACGGTGCCCTTGGCTGGGGAGGCTCTTGGTTCTG 409

QY  460 CCATCATGGCCAGATTATCAAGCTTGAAAGCCTCTCTTGACCTCTCTGACCCGACGTCACA 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  410 CCTTTTCATTGGCAACACGGTCGACATTTCTTCTTGGACCTCATATAATTCACCTCTCC 469

QY  520 ATTGAACATGATCCCAACACACACAGCCCTGTTCGCGTCTGGGAGAATGCCACTGAC 579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  470 ATTAAACCAATGCCAGCTGCGCTTCTCTGAGGTGACCATAGGCAAAATGCTCCCACTAAT 529

QY  580 CCAACCAAGCATCTCACTTTCCCTTCTCTGGACAAACACAGCAAAATCATCAAAAGTTCGTG 639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  530 ACAGATAAAATCTCCATTTCTCTGTGGGAGAGCATTACCCATCATCAACAGTATCTG 589

QY  640 AATAGCGTGATCAACAGCGCTGAAGACACTGTATCTCTCCCTGCTGCAGAGAGATATGT 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  590 GATGGTGCTCTACCCCTCTTTACAAGTACATGTCTCAAAGCTCTCTCAAACTTCCTATGT 649

QY  700 CCAGTGATCCGCATGTTTCATCCAGTCCCTGGATGTGAATGTCTATTCAAGAGTCTCGAT 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  650 CCATTGCTTCAGTACGCTCTCAGCACACTGAATCCAAGTGACTTCAGGGTCTCTCTCT 709

QY  760 AATCTCTCAGCACAAACCCAGCTGCCAACCCCTCATTTTGAAGAGGACGAATGAGGAGACC 819
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  710 AATCTACTGGCCAGGACGAGTAACAACATTTGCTCTGAAGAGGAAGACAGAGAGATGG 769

QY  800 ACTGTGGTGATGCTGATTGGTTCCAGTGGCTTGGCCCAACCCCTTATAGCATCTCCCT 879
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db	770	CACAACCTCTCTGGCTGTAAATCCAGGCTTTACACCTTACTCTGATGGCATTTCCTT	829
Qy	880	CCAGGAAGCTGCTGCACACCACTAAC	905
Db	830	CTACAAAGTGGCAACTTTTCATCACAC	855
RESULT	14		
LOCUS	BG865768		
DEFINITION	502784358F1 NCL1_CGAF_SG2 Mus musculus cDNA clone IMAGE:4910534	5'	EST 29-MAY-2001
ACCESSION	BG865768		
VERSION	1		
KEYWORDS	EST		
SOURCE	house mouse		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: Jeffrey E. Green, M.D.		
	cDNA Library Preparation: Life Technologies, Inc.		
	Mammalia Arrayed by: The I.M.A.G.E. Consortium (ILNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/ILNL at:		
	http://image.llnl.gov		
	Plate: LLNL0811	row: C	column: 15
	High quality sequence stop: 832.		
FEATURES	Location/Qualifiers		
source	1..1039		
	/organism="Mus musculus"		
	/strain="FVB/N"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:4910534"		
	/clone_lib="NCL_CGAP_SG2"		
	/lab_host="DH10B (TI phage-resistant)"		
	/note="Organ: salivary gland; Vector: pCMV-Sport6; Site:1;		
	Not1; Site:2; Sal1; Cloned unidirectionally. Primer: Oligo		
	dt. Average insert size 1.3 kb. Constructed by Life		
	Technologies. Note: this is a NCI_CGAP Library."		
BASE COUNT	271 a	268 c	205 g
ORIGIN			294 t
			1 others

Query Match	18.6%	Score 191.2	DB 10	Length 1039
Best local Similarity	55.1%	Pred. No. 3.7e-40		
Matches 505	Conservative	0	Mismatches 369	Indels 43
			Gaps	

QY	40	AGACAAAAGATCTTCACGCTTTGGAAACTTTGTTCTCTGTGGGGCGTGCTCACTGGGACC	99
Db	25	AAACCAAGAGTGTCCAACTTTGGAGCCTTGTGTCTTGTGTGGCTGCTCATGGGAAC	84
QY	100	TCAGAGTCTCTCTGTGACAATCTTTGGCAATGACCTTAAGCAATGTCGTGGATAAAGCTGGAA	159
Db	85	TCAGAGTCACATCTTTGGTGAAGTGTGGAGCGCTGTGAATAATTGA	130

QY	160	CCTGTCTTCTCAGAGGGACTTTGACACAGTTTGACAATACTCTTAAAGGCATCCTTGAGAAA	219
Db	131	-----AAATTTCTCAATCTCCCTCTGAAGCTGTCCCTCAGAAC	168
QY	220	CTGAAGGTGCACTAGGAGTGCTTCAGAAATCCAGTGCTTGGCACTGGCCAAAGCAGAG	279
Db	169	CTGAATCTGGATGTGGAGTGCTTTCAGCAGGCCAACAGTTGGCCATTTAGCCAAGAACAGC	228
QY	280	GCCCAGGAAGCTGAGAAATTCGTAACAATGTGCATTTCTAAAGCTGCTTCCAACATAACACG	339
Db	229	ATTCTAGAAAGCTTGAACACAGAGCGGACCTTGGCAATTTAAAAGCTTTACATCT	285


```

Qy 340 GACATTTTGGTTGAAATCAGCAACTCCCTCATCTCTGATGTCAAAGCTGAACGCAT 399
Db 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
286 AATGGTTATTGTTGAAATCAATATCTCAAGTCTCTGATTTCAAGCTTAAGCTGTCT 345
Qy 400 GATGATGCAAGCAAGCTTAACCTGAGCTTCCCTGTCTCACCAGCAATGTCACTGTGGCGGG 459
Db 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
346 TCCACAGCAAGCAAGCTTAACCTGAGCTTCCCTGTCTCACCAGCAATGTCACTGTGGCGGG 405
Qy 460 CCCATTTGGCCCAAGTATTCAACCTGAAAGCCTCTTGACCTCTGACCTGACCGAGTCACA 519
Db 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
406 CTTTTCATTTGCAAGAGCGGTGCAGATTTCTGTTTCTCTGACCTCATATAATTCACCTCTCC 465
Qy 520 ATTGAAGTATGATCCCAAGACACAGCAGCTGTGTCCTGCGCTGCGGGAATGCGCCAGTGAC 579
Db 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
466 ATTAACCAATGCCAGAGCTGCTTCTCTGAGTGAACATAGGCAAAATGCTCCAGTAAT 525
Qy 580 CCAACAGCATCTCACTTTCTCTGCTGCAACAAACACAGCCAAATCATCAACAGTTCGNG 639
Db 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
526 ACAGTAAATCTCCATTTCTTGTNGGAGAGAGATTACCCATCATCAACAGTATTCTG 585
Qy 640 AATAGCTGTATCAACAGCTGAAAGCACTGTATCTCTGCTGCTGCAAGAGAGATATGT 699
Db 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
586 GATGGTGTCTCTACCTCTCTTACAAGTACATTTGTCAACGCTCTGCAAAACTTCCTATGT 645
Qy 700 CCATGATCGCATCTTCACTCCACTCTCTGCTGATGTGAATGTCAATTCAGCAGGTCTGTCGAT 759
Db 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
646 CCATTTGCTTCACTAGTCTTCAACACTGAATCCAAAGTGTCTTCAAGTCTCTCTCTCT 705
Qy 760 AATCCTCAGCACAAACCCAGCTGCAACCCCTCATTTGAAGAGAGAGAA-TGAGGAGAGAC 818
Db 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
706 AATCTATTTGAGGAGTGT-ATTATATATCCACACATTTCTACTGAGACCAACCCAAAGAGAC 882
Qy 819 CACTGTGCTGATGCTGATTTGCTTCCAGTGGCTTCCGACCCCTTATAGCATCTCC 878
Db 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
764 CCAATCTTCTGCTGATCTCTCAGGGTTTCATCTTCTGATGGCATTTCCC 823
Qy 879 TCCAGGAGCTGTGTCACCACTCAACAGCGTGAAGCGCTGAGTCCCAACCAAGAGAC 938
Db 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
824 TCTTAGAAGTGT-ATTATATATCCACACATTTCTACTGAGACCAACCCAAAGAGAC 882
Qy 939 TTCCAGATACCCCTTC 955
Db 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
583 CTTCAGAAATCTCTTC 899

RESULT 15
BG868020
LOCUS 60278832F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4919756 5',
DEFINITION mRNA sequence.
ACCESSION BG868020
VERSION BG868020.1 GI:14218560
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 889)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10835 row: c column: 21
High quality sequence stop: 844.

```

```

FEATURES
source
Location/Qualifiers
1..889
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4919756"
/lab_host="NCI_CGAP_SG2"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site:1:
NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 Kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 235 a 235 c 169 g 251 t
ORIGIN
Query Match 18.6% Score 190.8; DB 10; Length 889;
Best Local Similarity 55.7%; Pred. No. 4; 3e-40;
Matches 428; Conservative 0; Mismatches 332; Indels 8; Gaps 3:
Qy 201 TAAAGGCATCCTTGAGAAACTGAAGTCGACCTAGGAGTCTTCAGAAATCCAGTGTCTG 260
Db 111 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
96 TGAAGCTGTCCCTCAGAACCTGAATCTGGATGTGGAGTTGCTTCAGCAGCCACAAGTTG 155
Qy 261 CCAACTGGCCAGCAGAGAGCCAGGAGCTGAGAAATTCCTGAACAATGTCTATTTCTAA 320
Db 111 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
156 GCCATTAGCCAAAGAACAGCATTTCTAGAAACGTTTGAACACAGCGGACCTTGGCAATTTAA 215
Qy 321 GCTGCTTCCAACTAACACGGACATTTTGGGTTGAAATCAGCAACTCCCTCATCTCTGGA 380
Db 111 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
216 AAGCTTTACATCTTTGAATGTTATT--GTTGAAATCAATATCTCAAGCTCTGGA 272
Qy 381 TGTCAAGCTGAACCGATCGATGATGCAAGGCTTTAACTGAGCTTCCTGTCTACCCG 440
Db 111 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
273 TTTTCAAGCTAAGCTGTCTTCCAAAGGCAATGGCATGTGATCTGACGGTGGCTTGGCTG 332
Qy 441 GAATGTCTACTGTGCGCGGCCATCATTTGCGCAGATTTATCAACTGAAAGCCCTCTTGA 500
Db 111 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
333 GGAAGCTCTTGGTCTGCTTTCATTTGCAAGAGCGTGCACATTTCTGTTTCTCTGGA 392
Qy 501 CTTCTGACCGCAGTCAATTTGAACTGATCCCGACAGACACAGCGCTGTGCCCTCTCT 560
Db 111 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
393 CCTCATAAATTCACCTCTCCATTAACCAATGCCCAGACTGGCTTCTCTGAGGTGACCAT 452
Qy 561 GGGAGATGGCCAGTGACCCAGCAGCATCTCACTTTCTCTGCTGGACAAACACAGCCA 620
Db 111 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
453 AGGCAATGCTCCAGTAATACAGATAAATCTCCATTTCTCTGTTGGAGAGCGATTACC 512
Qy 621 AATCATCAACAAAGTTCGTGAATAGCGTGTATCAACGCTGAAAGCACTGTATCTCTCT 680
Db 111 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
513 CATCATCAACAGTATTTCTGGATGGTGTCTCTACCTCTCTTACAAGTACATTTGCAACCT 572
Qy 681 GCTCAGAGGAGATATGTCACCTGATCCCGATCTTCATCCACTCTCTGATGTGATGT 740
Db 111 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
573 CTTCAAAACTTCTATGTCCTTGTCTAGTACGTCCTCAGCACACTGAATCCAAGTGT 532
Qy 741 CATTCAGAGTCTCTGATATATC--CTCAGCACAACCCAGCTGCAAACTCATTTGA 798
Db 111 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
633 TCTTCAGGCTCTCTCTCTAATCTACTTGCAGCAGCAGGTACAATTCGCCCTCTCAAGA 692
Qy 799 AGAGAGCAATGAGGAGGACCACTGTGGTGCATGCTGTTGGTTCCAGTGGCTTCCGCC 858
Db 111 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
693 GGAAGAACCAAGAGGATGCCAATAATTCCTCTGGGTACTCTCCAGTGGTGTTCATCT 752
Qy 859 ACCCCCTTATAGCATCTCCCTCCAGGAGGCTGCTGCCACCACTTAACACAGCGTGAAGC 918
Db 111 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
753 TACTTCTGATGGCATTTCCCTCTAGAAAAAGTGT---ACTAATCATCCAGAAATTTCTACC 809
Qy 919 TGAGTCCCAAGAGGACCTTCCAGATACCCCTTCTCTCTCTCAGATC 966
Db 111 111 1111111111 11 1111111111 1111111111 1111111111 1111111111
810 TGAGAACCAACCAAGGACCTCCAAAATTTCTCTCTCTCTCTGAGTAGTC 857

```

Search completed: August 6, 2002, 16:52:39
Job time: 10146 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2002, 16:18:33 : Search time 208.17 seconds
(without alignments)
8478.590 Million cell updates/sec

Title: US-10-020-139-1

Perfect score: 1028

Sequence: 1 CACGAGATTTCATGACATC.....AAGGCCCTTTCTGCAAAA 1028

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

1:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDSI/qcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1028	100.0	1028	19 AAV44759	Human parotid secr
2	1019.8	99.2	1058	21 AAB7727	Human secreted pro
3	1019.8	99.2	1058	22 AAF54009	cDNA encoding huma
4	994.4	96.7	1041	22 AAH59686	Human protein enco
5	990.4	96.3	1049	21 AAC58379	Human PRO1025 nucl
6	818.4	79.6	824	19 AAV37699	Human parotid secr
7	383.8	37.3	449	19 AAV44760	Expressed sequence
8	374.8	36.5	538	19 AAV44761	Expressed sequence
9	312.8	30.4	359	19 AAV44762	Expressed sequence

10	303.6	29.5	374	19 AAV44763	Expressed sequence
11	302.4	29.4	317	21 AAC23485	Human secreted pro
12	258.8	25.2	406	19 AAV44764	Expressed sequence
13	229.8	22.4	493	19 AAV44765	Expressed sequence
14	191	18.6	653	21 AAA61274	Human secreted pro
15	173.4	16.9	7524	22 AAK89798	Human digestive sy
16	165.2	16.1	581	22 ABA60313	Human foetal liver
17	165.2	16.1	581	22 AAK08593	Human brain expres
18	165.2	16.1	581	22 AAK34475	Human bone marrow
19	165.2	16.1	581	22 AAI40197	Probe 8883 used t
20	153	14.9	153	22 ABA72852	Human foetal liver
21	153	14.9	153	22 AAK21285	Human brain expre
22	153	14.9	153	22 AAK47441	Human bone marrow
23	153	14.9	153	22 AAI53276	Probe #21962 used
24	129	12.5	395	19 AAV44766	Expressed sequence
25	103.2	10.0	116	19 AAV44767	Expressed sequence
26	42.2	4.1	800	22 AAK87987	Human digestive sy
27	41.5	4.0	360	19 AAV44768	Expressed sequence
28	40.4	3.9	768	22 AAH27756	Human LUNX cDNA.
29	40.4	3.9	1009	20 AAX00808	LS170 gene derived
30	40.4	3.9	1015	22 AAH27757	Human LUNX cDNA.
31	40.4	3.9	1016	21 AAZ61165	Nucleotide sequenc
32	40.4	3.9	1017	20 AAX00809	LS170 gene consens
33	40.4	3.9	1021	22 AAS46111	Human DNA encoding
34	40.4	3.9	1060	20 AAX59356	Human secreted pro
35	40.4	3.9	1060	22 AAS59237	Human cDNA encodin
36	40.4	3.9	1060	24 ABA90906	Human polynucleoti
37	40.4	3.9	1071	22 AAI58877	Human polynucleoti
38	40.4	3.9	1084	22 AAI60663	Human polynucleoti
39	40.4	3.9	1125	22 AAK87982	Human digestive sy
40	40	3.9	840	22 AAK87985	Human digestive sy
41	39	3.8	3395	22 AAI59757	Human polynucleoti
42	39	3.8	3423	22 AAI57971	Human polynucleoti
43	39	3.8	4062	22 AAD05509	Human polynucleoti
44	38.6	3.8	429	19 AAV43243	MSRV-1 clone 8 pol
45	38.6	3.8	837	22 AAH68310	C glutamicum codin

ALIGNMENTS

RESULT 1
AAV44759
ID AAV44759 standard; DNA; 1028 Bp.
XX AC AAV44759;
XX DT 16-OCT-1998 (first entry)
XX DE Human parotid secretory protein coding sequence.
XX KW Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
XX KW non-immune defensive disorder; immune system disorder; cancer; human;
XX KW therapy; diagnosis; ss.
XX OS Homo sapiens.
XX FH Key
XX CDS Location/Qualifiers
FT 49..798
FT /*tag= a
FT /product= hPSP
FT sig_peptide 49..102
FT /*tag= b
FT mat_peptide 103..795
FT /*tag= c

XX WO9828420-A1.
XX PD 02-JUL-1998.
XX PF 18-DEC-1997; 97WO-US23522.
XX PR 23-DEC-1996; 96US-0034429.

Qy 181 GAGACAGTTGACATCTCTTAAAGG-----CATCTTGGAGAACTGAAGGTGGACCT 233
 |||||
 Db 183 gagacagttgacatactctaaaggccatcccnlttngagaaactgaaggtcgacct 242
 |||||
 Qy 234 AGGAGTCTCTCAGAAATCCAGTCTTGGCAACTGGCCAAAGCAGAAAGGCCAGGAAGCTGA 293
 |||||
 Db 243 aggagttctcaaaatccagttgcttggcaactggccaaagcagaagggcccaagaaactga 302
 |||||
 Qy 294 GAAATTCCTGAA--CAATGTCTATTTCTAAAGCTGCTTCCAACTAACACGGACATTTTGGGT 352
 |||||
 Db 303 qaaattgctgaacaaatgctatttctlaagctgcttccaaactaaacgggacattttgggg 362
 |||||
 Qy 353 TG-AAATCAGCAACTCCCTCATCTCTGGGATGTCAAAGCTTGAACCCGATCGATGGCRAA 411
 |||||
 Db 363 tgaataatcagcaactccctcatcttggatgtcaaaagctgaaccggatgatggnaaa 422
 |||||
 Qy 412 GGCCTTAACCTGAGCTTCCCTGTCAACC 438
 |||||
 Db 423 ggctaaactggggtctccctgtcanc 449

RESULT 8

AAV44761
 ID AAV44761 standard; DNA; 538 BP.

AC AAV44761;

DT 16-OCT-1998 (first entry)

DE Expressed sequence tag HSGSCL3R.

KW Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
 non-immune defensive disorder; immune system disorder; cancer; human;
 therapy; diagnosis; ss.

OS Homo sapiens.

XX WO9828420-A1.

XX 02-JUL-1998.

XX 18-DEC-1997; 97WO-US23522.

XX 23-DEC-1996; 96US-0034429.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Duan R, Ruben SM;

XX WPI; 1998-377651/32.

PT New nucleic acid encoding human parotid secretory protein or its
 fragments - useful for diagnosis and treatment of, e.g. digestive
 and endocrine disorders and for drug screening

PS Disclosure; Page 76; 94pp; English.

XX This sequence represents an expressed sequence tag with homology to the
 CC DNA of the invention, and is all so specifically stated as not being
 CC contained within the DNA of the invention. The DNA of the invention
 CC encodes the human parotid secretory protein (hPSP). The hPSP DNA is
 CC useful for chromosome identification and isolation of the corresponding
 CC genomic DNA. The DNA and protein can be used to detect abnormal levels of
 CC hPSP (in standard blotting, amplification or immuno assays), particularly
 CC for diagnosis of digestive, non-immune defensive, endocrine or immune
 CC system disorders. A particular application is diagnosis of cancers of the
 CC salivary gland, thymus and pancreas which are associated with high levels
 CC of hPSP. The protein is also useful as antifungal, antibacterial,
 CC antiparasitic and antiviral agents and may be expressed in vivo from the
 CC DNA. The protein, or cells expressing it, are used in screening tests to
 CC identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and
 CC antisense nucleic acids, which are potentially useful for treating
 CC conditions associated with excessive hPSP production. Cells containing

CC the DNA are used to express the recombinant protein and this can be used
 CC to raise Ab, useful for diagnosis, therapy, for affinity purification and
 CC to identify hPSP-binding proteins.

XX Sequence 538 BP: 140 A; 126 C; 128 G; 124 T; 20 other;

Query Match 36.5%; Score 374.8; DB 19; Length 538;
 Best Local Similarity 87.7%; Pred. No. 2.5e-103;
 Matches 469; Conservative 0; Mismatches 54; Indels 12; Gaps 6;

Qy 12 ATGACGATCTCTCTTAACGCGCTGTCAAGACAAAGATGCT-TCAGCTTTGGAACCTTG 70
 |||||
 Db 1 atgagcatctctctaaacgctgtcaagacaaagatgctnncagcttggaaacttg 60
 |||||
 Qy 71 TTCTCTCTGTCGGCGTCTCACTGGGACCTTCAGAGTCTCTTCTTGACAATCTTGGCAATG 130
 |||||
 Db 61 tctctctgtggcggtgctcaetgggaacctcagagctctcttcttgacaatcttggcaatg 120
 |||||
 Qy 131 ACCTAACCAATGTCGTGGATTAAGCTGGAACCTTCTTTCACGAGGACTTGACAGACTTG 190
 |||||
 Db 121 acctaaagcaatgtctgtgataaagctggaacctgtcttcacaggagacttgagacagttg 180
 |||||
 Qy 191 ACAATACTCTTAAAGGCACTCTTGAGAAACTGAAGCTCGACCTAGGAGTCTTCAGAAAT 250
 |||||
 Db 181 acaatactctaaagcactcttgagaaactgaagctcagactagagagtgcttcagaaat 240
 |||||
 Qy 251 CCAGTCTTGGCAACTGGCCAAAGCAGACAGGCCAGAACTGAGAAATTCGTGAACAATG 310
 |||||
 Db 241 ccagtgcttggaactggccaa--cagaaggnccaggaaagctgagaaattgctgaaactg 299
 |||||
 Qy 311 TCATTTCTAAGCTGCTTCCAACTAACACGACATTTTGGGTGAAATCAGCAACTCCC 370
 |||||
 Db 300 tcatcttaagctgcttccaaactaacacaggaacttttgggttgaaantcagcaatnccc 359
 |||||
 Qy 371 TCATCTCTGATGTCAAGCTGAACCGATCGATGCGCAAGGCCCTTAACC---TGAGC 426
 |||||
 Db 360 ncanccgagtggttcaaaagntgnaucgatgatgagtggaaggtcttaancgcggaggt 419
 |||||
 Qy 427 TTCCTGTCAACCGGGAATGT--CACTGTGGCGGGGCCCATCATTT--GGCGAGATTCAA 482
 |||||
 Db 420 tccgtgtcaccgggaatgttcaangtngccggccctctatg99ccagnttatcaaa 479
 |||||
 Qy 483 --CCTGAACCCCTCTTGGACCTCTGACCGCAGTCACAAATGAAACTGATCCCC 535
 |||||
 Db 480 anctggaaagcttccctgggacctccggacngntcaacaattgaaangattcccc 534
 |||||

RESULT 9

AAV44762

XX ID AAV44762 standard; DNA; 359 BP.

AC AAV44762;

DT 16-OCT-1998 (first entry)

DE Expressed sequence tag HSGSA89R.

XX Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
 KW non-immune defensive disorder; immune system disorder; cancer; human;
 KW therapy; diagnosis; ss.

OS Homo sapiens.

XX WO9828420-A1.

XX 02-JUL-1998.

XX 18-DEC-1997; 97WO-US23522.

XX 23-DEC-1996; 96US-0034429.

XX (HUMA-) HUMAN GENOME SCI INC.

XX PI Duan R, Ruben SM;
XX KW
XX KW
XX OS Homo sapiens.
XX PN WO9828420-A1.
XX PD 02-JUL-1998.
XX PF 18-DEC-1997; 97WO-US23522.
XX PR 23-DEC-1996; 96US-0034429.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Duan R, Ruben SM;
XX WI 1998-377651/32.
XX CC New nucleic acid encoding human parotid secretory protein or its fragments - useful for diagnosis and treatment of, e.g. digestive and endocrine disorders and for drug screening
XX PS Disclosure; Page 77; 94pp; English.
XX CC This sequence represents an expressed sequence tag with homology to the DNA of the invention, and is all so specifically stated as not being contained within the DNA of the invention. The DNA of the invention encodes the human parotid secretory protein (hPSP). The hPSP DNA is useful for chromosome identification and isolation of the corresponding genomic DNA. The DNA and protein can be used to detect abnormal levels of hPSP (in standard blotting, amplification or immuno assays), particularly for diagnosis of digestive, non-immune defensive, endocrine or immune system disorders. A particular application is diagnosis of cancers of the salivary gland, thymus and pancreas which are associated with high levels of hPSP. The protein is also useful as antifungal, antibacterial, antiparasitic and antiviral agents and may be expressed in vivo from the DNA. The protein, or cells expressing it, are used in screening tests to identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and antisense nucleic acids, which are potentially useful for treating conditions associated with excessive hPSP production. Cells containing the DNA are used to express the recombinant protein and this can be used to raise Ab, useful for diagnosis, therapy, for affinity purification and to identify hPSP-binding proteins.
XX SQ Sequence 359 BP; 94 A; 79 C; 90 G; 90 T; 6 other;

Query Match 30.4%; Score 312.8; DB 19; Length 359;
Best Local Similarity 97.2%; Pred. No. 1.3e-84;
Matches 347; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

QY 1 CAGGAGATTTCATGACATCTCTCTTAACCGGTGTCAGACAAAGATGCTTCAGCTT 60
DB 3 cagagatttcgatgacatctctcttaaacggtgtcagacaaagatgcttcagctt 62
QY 61 TGAAGACTTGTCTCTGTGGGGGTGCTCACTGGGACCTCAGAGTCTCTTCATGACAAAT 120
DB 63 tgaagacttgtctctgtgggggtgctcactgggacctcagagctctcttcagacaaat 122
QY 121 CTGTGCAATGACCTAAGCAATGCTGTGGATAAGCT-GGAACCTGTCTTTCACGAGGACT 179
DB 123 ctgtgcaatgacctaagcaatgctgtggataagct-ggaacctgtcttcacgaggact 182
QY 180 TGAGACAGTTGACATACATCTTAAAGGACATCTTGAAGAACTGAGCTGACCTAGGAGT 239
DB 183 tgagacagttgacatacatctttaaaggacatcttgagaaactgagctgacctaggagt 242
QY 240 GCTTTCAGAAATCCAGTGTGTCGAACCTGGCCAAAGCAGAGG-CCCAAGAACTGAGAAAT 298
DB 243 gctttcagaaatccagtgtgtgcaactggncaagcagaaggnccagagagctgagaaat 302
QY 299 TGCT-GAACAATGTCTATTTCTAAGCTGCTTCCAACTAACACGGACATTTTTCGGTTG 354
DB 303 tgcctgancatgtcaattcttaagctgnttccgactaacacacgncatttttgggttg 359

RESULT 10
ID AAV44763
XX AAV44763 standard; DNA; 374 BP.
XX AC AAV44763;
XX AC
XX DT 16-OCT-1998 (first entry)
XX DT
XX DE Expressed sequence tag HSPA114R.
XX DE

Parotid secretory protein; hPSP; digestive disorder; endocrine disorder; non-immune defensive disorder; immune system disorder; cancer; human; therapy; diagnosis; ss.
Homo sapiens.
WO9828420-A1.
02-JUL-1998.
18-DEC-1997; 97WO-US23522.
23-DEC-1996; 96US-0034429.
(HUMA-) HUMAN GENOME SCI INC.
Duan R, Ruben SM;
WPI; 1998-377651/32.
New nucleic acid encoding human parotid secretory protein or its fragments - useful for diagnosis and treatment of, e.g. digestive and endocrine disorders and for drug screening
Disclosure; Page 77; 94pp; English.
This sequence represents an expressed sequence tag with homology to the DNA of the invention, and is all so specifically stated as not being contained within the DNA of the invention. The DNA of the invention encodes the human parotid secretory protein (hPSP). The hPSP DNA is useful for chromosome identification and isolation of the corresponding genomic DNA. The DNA and protein can be used to detect abnormal levels of hPSP (in standard blotting, amplification or immuno assays), particularly for diagnosis of digestive, non-immune defensive, endocrine or immune system disorders. A particular application is diagnosis of cancers of the salivary gland, thymus and pancreas which are associated with high levels of hPSP. The protein is also useful as antifungal, antibacterial, antiparasitic and antiviral agents and may be expressed in vivo from the DNA. The protein, or cells expressing it, are used in screening tests to identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and antisense nucleic acids, which are potentially useful for treating conditions associated with excessive hPSP production. Cells containing the DNA are used to express the recombinant protein and this can be used to raise Ab, useful for diagnosis, therapy, for affinity purification and to identify hPSP-binding proteins.
Sequence 374 BP; 86 A; 123 C; 82 G; 78 T; 5 other;

Query Match 29.5%; Score 303.6; DB 19; Length 374;
Best Local Similarity 95.0%; Pred. No. 8.3e-82;
Matches 343; Conservative 0; Mismatches 14; Indels 4; Gaps 3;

QY 644 GCCTGATCAACACGCTGAAAGCACTGTATCTCTCTGCTGAGAGGAGATATCTCCAC 703
DB 14 gcctgatcaacacgctgaaagcactgtatctctctgctgagagagagatattctccac 73
QY 704 TGATCGCATCTTCATCCACTCCCTGGATGTGAATGTTCATTTCAGCAGGTGCTGCGATAATC 763
DB 74 tgatcgcatcttcatccactccctggatgtgaatgttcatttcagcaggtgctgcgataatc 133
QY 764 CTCAGCACAAAACCCAGCTGCAAACTTCATTGTAAGAGGACGAATGAGGAGGACCACTG 823
DB 134 ctacagcacaacccagctgcaaaacttcattgttaagaggacgaatgaggaggaccactg 193
QY 824 TGGTGCATGCTGATTGGTTCCAGTGGCTGCCACCCCTTATAGCATCTCCCTCCAG 883
DB 194 tggatgcattgattggttccagtggttccaccccttattagcatctccctccag 253
QY 884 GAAGCTGTGCCACACACTAACACCGTGAAGAGCT-GAGTCCCAAGAGGAGGACCTTCC 942
DB 254 gaagctgtgccacacactaacacggtgaagagct-gagtcccaagaggagaccttcc 313

XX SQ Sequence 406 BP; 110 A; 95 C; 95 G; 100 T; 6 other;

Query Match 25.2%; Score 258.8; DB 19; Length 406;
Best Local Similarity 89.4%; Pred. No. 3.4e-68;
Matches 313; Conservative 0; Mismatches 29; Indels 8; Gaps 3;

QY 9 TTCATCAGCATCCTCCTCTAAGCGCTCAAGACAAAGATGCTTCAGCTTTGCAACT 68
DB 1 TCAAGAGCATCCTCCTCTAAGCGCTCAAGACAAAGATGCTTCAGCTTTGCAACT 59
QY 69 TGTCTCTCTGTCGCGCTGCTCACTGGACCTCAGAGTCTCTTCTTGACAATCTTGCAA 128
DB 60 TGTCTCTCTGTCGCGCTGCTCACTGGACCTCAGAGTCTCTTCTTGACAATCTTGCAA 119
QY 129 TGACCTAAGCAATGCTCGTGGATAGCTGGAACCTCTCTTTCACGAGGACCTTGAGACAGT 188
DB 120 TGACCTAAGCAATGCTCGTGGATAGCTGGAACCTCTCTTTCACGAGGACCTTGAGACAGT 179
QY 189 TGACATACTCTTAAAGGCATCCTTTGAGAACTGAAGTGCACCTAGACGTGCTTCAGAA 248
DB 180 TGACATACTCTTAAAGGCATCCTTTGAGAACTGAAGTGCACCTAGACGTGCTTCAGAA 239
QY 249 ATCCAGTGTTCGCAACTGGCCCAACGACAGAGGCC---AGGAAGCTGAGAAATTCGTG-- 303
DB 240 ATCCAGTGTTCGCAACTGGCCCAACGAGAGAGGCCAGAGGAGAGAGAAATTCGTG-- 299
QY 304 --AACAAATGCTATTTCTAAGCTGCTTCCAACTAACACGGACATTTTGGG 351
DB 300 ACAAATGCTATTTCTAAGCTGCTTCCAACTAACACGGAGGACATTTTGGG 349

RESULT 13
AAV44765
ID AAV44765 standard; DNA; 493 BP.
AC AAV44765;
XX
DT 16-OCT-1998 (first entry)
XX
DE Expressed sequence tag HSPMD56R.
XX
KW Parotid secretory protein; hsp; digestive disorder; endocrine disorder;
KW non-immune defensive disorder; immune system disorder; cancer; human;
KW therapy; diagnosis; ss.
XX
OS Homo sapiens.
XX
PN WO9828420-A1.
XX
PD 02-JUL-1998.
XX
PF 18-DEC-1997; 97WO-US23522.
XX
PR 23-DEC-1996; 96US-0034429.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Duan R, Ruben SM;
XX
WPI; 1998-377651/32.
XX
PT New nucleic acid encoding human parotid secretory protein or its
PT fragments - useful for diagnosis and treatment of, e.g. digestive
PT and endocrine disorders and for drug screening
XX
PS Disclosure; Page 78; 94pp; English.
XX
CC This sequence represents an expressed sequence tag with homology to the
CC DNA of the invention, and is all so specifically stated as not being
CC contained within the DNA of the invention. The DNA of the invention
CC encodes the human parotid secretory protein (hPSP). The hPSP DNA is

CC useful for chromosome identification and isolation of the corresponding
CC genomic DNA. The DNA and protein can be used to detect abnormal levels of
CC hPSP (in standard blotting, amplification or immuno assays), particularly
CC for diagnosis of digestive, non-immune defensive, endocrine or immune
CC system disorders. A particular application is diagnosis of cancers of the
CC salivary gland, thymus and pancreas which are associated with high levels
CC of hPSP. The protein is also useful as antifungal, antibacterial,
CC antiparasitic and antiviral agents and may be expressed in vivo from the
CC DNA. The protein, or cells expressing it, are used in screening tests to
CC identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and
CC antisense nucleic acids, which are potentially useful for treating
CC conditions associated with excessive hPSP production. Cells containing
CC the DNA are used to express the recombinant protein and this can be used
CC to raise Ab, useful for diagnosis, therapy, for affinity purification and
CC to identify hPSP-binding proteins.
XX
SQ Sequence 493 BP; 118 A; 143 C; 107 G; 122 T; 3 other;

Query Match 22.4%; Score 229.8; DB 19; Length 493;
Best Local Similarity 93.4%; Pred. No. 2.3e-59;
Matches 240; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 602 TCCTGACAAACACAGCCCAATCATCAACAGTTCTCGTAAGCTGATCAACAGCTGA 661
DB 83 TGTTCCTGACAAACACAGCCCAATCATCAACAGTTCTCGTAAGCTGATCAACAGCTGA 142
QY 662 AAAGCACTGTATCTCTCCCTGCTGCAGAGGAGATGTCCACTGATCCGATCTTTCATCC 721
DB 143 AAAGCACTGTATCTCTCCCTGCTGCAGAGGAGATGTCCACTGATCCGATCTTTCATCC 202
QY 722 ACTCCCTGGATGTGAATGTCTATTCAGCAGGCTGCTCGATTAATCTCAGCACAACCCAGC 781
DB 203 ACTCCCTGGATGTGAATGTCTATTCAGCAGGCTGCTCGATTAATCTCAGCACAACCCAGC 262
QY 782 TGCAAAACCTCATTTGAAGAGGACGAATGAGGAGGACCACTGTGTGCTGATGTTGGT 841
DB 263 TGCAAAACCTCATTTGAAGAGGACGAATGAGGAGGACCACTGTGTGCTGATGTTGGT 322
QY 842 TCCAGTGGCTTGCCTCC 858
DB 323 AGCCAGTCTCTGTGCC 339

RESULT 14
AA61274
ID AAA61274 standard; DNA; 653 BP.
AC AAA61274;
XX
DT 18-OCT-2000 (first entry)
XX
DE Human secreted protein gene 15 clone HSPBY40.
XX
KW Human; secreted protein; fusion protein; gene therapy;
KW protein therapy; diagnosis; tissue; cancer; tumour; AIDS;
KW autoimmune disorder; allergy; cardiovascular; viral; bacterial;
KW fungal infection; immunosuppressive; ds.
XX
OS Homo sapiens.
XX
PN WO200029422-A1.
XX
PD 25-MAY-2000.
XX
PF 09-NOV-1999; 99WO-US26409.
XX
PR 12-NOV-1998; 98US-0108207.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J, Ruben SM, Rosen CA, Ebner R, Florence KA, Young PE;
PI Birse CE, Carter KC, Komatsoulis G;

PR	31-JAN-2000;	2000US-0179062;
PR	04-FEB-2000;	2000US-0180621;
PR	24-FEB-2000;	2000US-0184664;
PR	02-MAR-2000;	2000US-0186550;
PR	16-MAR-2000;	2000US-0189874;
PR	17-APR-2000;	2000US-0190176;
PR	18-APR-2000;	2000US-0191823;
PR	19-MAY-2000;	2000US-0205515;
PR	07-JUN-2000;	2000US-0209467;
PR	28-JUN-2000;	2000US-0214886;
PR	30-JUN-2000;	2000US-0215135;
PR	07-JUL-2000;	2000US-0216647;
PR	07-JUL-2000;	2000US-0216880;
PR	11-JUL-2000;	2000US-0217487;
PR	11-JUL-2000;	2000US-0217496;
PR	14-JUL-2000;	2000US-0218290;
PR	14-JUL-2000;	2000US-0218290;
PR	26-JUL-2000;	2000US-0220963;
PR	26-JUL-2000;	2000US-0220964;
PR	14-AUG-2000;	2000US-0224511;
PR	14-AUG-2000;	2000US-0224511;
PR	14-AUG-2000;	2000US-0225213;
PR	14-AUG-2000;	2000US-0225214;
PR	14-AUG-2000;	2000US-0225266;
PR	14-AUG-2000;	2000US-0225267;
PR	14-AUG-2000;	2000US-0225268;
PR	14-AUG-2000;	2000US-0225621;
PR	14-AUG-2000;	2000US-0225627;
PR	22-AUG-2000;	2000US-0226681;
PR	22-AUG-2000;	2000US-0226868;
PR	22-AUG-2000;	2000US-0227182;
PR	23-AUG-2000;	2000US-0227009;
PR	30-AUG-2000;	2000US-0228924;
PR	01-SEP-2000;	2000US-0229288;
PR	01-SEP-2000;	2000US-0229343;
PR	01-SEP-2000;	2000US-0229344;
PR	01-SEP-2000;	2000US-0229345;
PR	05-SEP-2000;	2000US-0229509;
PR	05-SEP-2000;	2000US-0229513;
PR	06-SEP-2000;	2000US-0230437;
PR	08-SEP-2000;	2000US-0230438;
PR	08-SEP-2000;	2000US-0231242;
PR	08-SEP-2000;	2000US-0231243;
PR	08-SEP-2000;	2000US-0231244;
PR	08-SEP-2000;	2000US-0231413;
PR	08-SEP-2000;	2000US-0232401;
PR	08-SEP-2000;	2000US-0232414;
PR	08-SEP-2000;	2000US-0232080;
PR	08-SEP-2000;	2000US-0232081;
PR	12-SEP-2000;	2000US-0233968;
PR	14-SEP-2000;	2000US-0233997;
PR	14-SEP-2000;	2000US-0233998;
PR	14-SEP-2000;	2000US-0233999;
PR	14-SEP-2000;	2000US-0234000;
PR	14-SEP-2000;	2000US-0234011;
PR	14-SEP-2000;	2000US-0233063;
PR	14-SEP-2000;	2000US-0233064;
PR	14-SEP-2000;	2000US-0233065;
PR	21-SEP-2000;	2000US-0234222;
PR	21-SEP-2000;	2000US-0234274;
PR	25-SEP-2000;	2000US-0234997;
PR	25-SEP-2000;	2000US-0234998;
PR	26-SEP-2000;	2000US-0235834;
PR	27-SEP-2000;	2000US-0235834;
PR	29-SEP-2000;	2000US-0235836;
PR	29-SEP-2000;	2000US-0236367;
PR	29-SEP-2000;	2000US-0236368;
PR	29-SEP-2000;	2000US-0236369;
PR	02-OCT-2000;	2000US-0236602;

xx

PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0245474.
PR 08-NOV-2000; 2000US-0245475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251888.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-502630/55.

Polynucleotides encoding digestive system antigens, useful for
diagnosing, treating, preventing and/or prognosing disorders of the
digestive system, particularly cancer and cancer metastases -

XX
PS
XX

Disclosure; SEQ ID NO 3374; 985pp; English.

CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.

XX
SQ Sequence 7524 BP; 2170 A; 1586 C; 1582 G; 2186 T; 0 other;

Query Match 16.9%; Score 173.4; DB 22; Length 7524;

Best Local Similarity 99.4%; Pred. No. 1.3e-41;

Matches 174; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 34 GTCTCAAGACAAAGATGCTTCAGCTTTGGAACTTGTTCTCTGTGGCGGCTGTCTACT 93
|||||Db 963 gtgtcaagacaaaagatgtctcagcttggaaactgttctctgtggtggtgctact 1022
|||||QY 94 GGGACCTCAGAGTCTCTTCTTGACATCTTGGCAATGACCTAAGCAATCTCTGGGATAG 153
|||||Db 1023 gggacctcagagctctctcttggcaacttggcaatgacctgaagcaatgtgtgataag 1082
|||||QY 154 CTGGAACCTGTTCTTCACGAGGACTTGACAGTTGACAATACTCTTAAAGGCA 208
|||||Db 1083 ctggaaacctgttcttcaagaggacttgagacagttgacaatactcttaaggta 1137
|||||Search completed: August 6, 2002, 17:28:16
Job time: 4183 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: August 6, 2002, 16:25:18 : Search time 2876.68 Seconds
(without alignments)
7731.369 Million cell updates/sec

Title: US-10-020-139-1
Perfect score: 1028
Sequence: 1 CACGAGATTTCATGAGCATC.....AAGGCCATTTCGCAAAA 1028

Scoring table: IDENTITY_NUC
Gapex 10.0, Gapext 1.0

Searched: 21979536 seqs, 1081749327 residues

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

- 1: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
- 2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
- 3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
- 4: /cgn2_6/ptodata/1/pna/US08_COMB.seq:*
- 5: /cgn2_6/ptodata/1/pna/US081_COMB.seq:*
- 6: /cgn2_6/ptodata/1/pna/US082_COMB.seq:*
- 7: /cgn2_6/ptodata/1/pna/US083_COMB.seq:*
- 8: /cgn2_6/ptodata/1/pna/US084_COMB.seq:*
- 9: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
- 10: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
- 11: /cgn2_6/ptodata/1/pna/US087_COMB.seq:*
- 12: /cgn2_6/ptodata/1/pna/US088_COMB.seq:*
- 13: /cgn2_6/ptodata/1/pna/US089_COMB.seq:*
- 14: /cgn2_6/ptodata/1/pna/US090_COMB.seq:*
- 15: /cgn2_6/ptodata/1/pna/US091_COMB.seq:*
- 16: /cgn2_6/ptodata/1/pna/US092_COMB.seq:*
- 17: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*
- 18: /cgn2_6/ptodata/1/pna/US094_COMB.seq:*
- 19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq:*
- 20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq:*
- 21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq:*
- 22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq:*
- 23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq:*
- 24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq:*
- 25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq:*
- 26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq:*
- 27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq:*
- 28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*
- 29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq:*
- 30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq:*
- 31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq:*
- 32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq:*
- 33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq:*
- 34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq:*
- 35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq:*
- 36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq:*
- 37: /cgn2_6/ptodata/1/pna/US100_COMB.seq:*
- 38: /cgn2_6/ptodata/1/pna/US101_COMB.seq:*
- 39: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*
- 40: /cgn2_6/ptodata/1/pna/US6001_COMB.seq:*
- 41: /cgn2_6/ptodata/1/pna/US6002_COMB.seq:*
- 42: /cgn2_6/ptodata/1/pna/US6003_COMB.seq:*
- 43: /cgn2_6/ptodata/1/pna/US6004_COMB.seq:*

44: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*

45: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*

46: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*

47: /cgn2_6/ptodata/1/pna/US6008_COMB.seq:*

48: /cgn2_6/ptodata/1/pna/US6009_COMB.seq:*

49: /cgn2_6/ptodata/1/pna/US6010_COMB.seq:*

50: /cgn2_6/ptodata/1/pna/US6011_COMB.seq:*

51: /cgn2_6/ptodata/1/pna/US6012_COMB.seq:*

52: /cgn2_6/ptodata/1/pna/US6013_COMB.seq:*

53: /cgn2_6/ptodata/1/pna/US6014_COMB.seq:*

54: /cgn2_6/ptodata/1/pna/US6015_COMB.seq:*

55: /cgn2_6/ptodata/1/pna/US6016_COMB.seq:*

56: /cgn2_6/ptodata/1/pna/US6017_COMB.seq:*

57: /cgn2_6/ptodata/1/pna/US6018_COMB.seq:*

58: /cgn2_6/ptodata/1/pna/US6019_COMB.seq:*

59: /cgn2_6/ptodata/1/pna/US6020_COMB.seq:*

60: /cgn2_6/ptodata/1/pna/US6021_COMB.seq:*

61: /cgn2_6/ptodata/1/pna/US6022_COMB.seq:*

62: /cgn2_6/ptodata/1/pna/US6023_COMB.seq:*

63: /cgn2_6/ptodata/1/pna/US6024_COMB.seq:*

64: /cgn2_6/ptodata/1/pna/US6025_COMB.seq:*

65: /cgn2_6/ptodata/1/pna/US6026_COMB.seq:*

66: /cgn2_6/ptodata/1/pna/US6027_COMB.seq:*

67: /cgn2_6/ptodata/1/pna/US6028_COMB.seq:*

68: /cgn2_6/ptodata/1/pna/US6029_COMB.seq:*

69: /cgn2_6/ptodata/1/pna/US6030_COMB.seq:*

70: /cgn2_6/ptodata/1/pna/US6031_COMB.seq:*

71: /cgn2_6/ptodata/1/pna/US6032_COMB.seq:*

72: /cgn2_6/ptodata/1/pna/US6033_COMB.seq:*

73: /cgn2_6/ptodata/1/pna/US6034_COMB.seq:*

74: /cgn2_6/ptodata/1/pna/US6035_COMB.seq:*

75: /cgn2_6/ptodata/1/pna/US6036_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1028	100.0	1028	13	US-08-993-529-1
2	1019.8	99.2	1058	18	US-09-469-099-26
3	1019.8	99.2	1058	22	US-09-598-360B-26
4	1018.4	99.1	1057	29	US-09-758-472-3794
5	1018.2	99.0	1122	25	US-09-652-122-4943
6	1018.2	99.0	1122	29	US-09-726-170-2263
7	1018.2	99.0	1122	29	US-09-726-175-2853
8	994.4	96.7	1041	1	PCT-US00-35017A-521
9	994.4	96.7	1041	18	US-09-471-275-3095
10	990.4	96.3	1049	37	US-10-081-056-163
11	981.4	93.5	1106	71	US-60-324-185-33384
12	818.4	79.5	824	11	US-08-749-288-2
13	631	61.4	643	25	US-09-652-122-2764
14	499.4	48.6	501	64	US-60-250-830-2204
15	499.4	48.6	501	71	US-60-323-966-2204
16	494.6	48.1	653	29	US-09-726-170-1233
17	490.4	47.7	531	32	US-09-834-366-5582
18	490.4	47.7	531	58	US-60-187-873-5582
19	473.2	46.0	631	29	US-09-758-472-173
20	462.6	45.0	477	32	US-09-834-366-46041
21	462.6	45.0	477	58	US-60-197-873-46041
22	457.8	44.5	507	29	US-09-726-170-1950
23	442.8	43.1	479	25	US-09-652-122-2930
24	435	42.3	460	29	US-09-726-170-1741
25	414.2	40.3	495	19	US-09-528-409-95211
26	414.2	40.3	495	35	US-09-933-524-95211
27	414.2	40.3	495	35	US-09-933-524A-95211
28	414	40.3	424	19	US-09-528-409-97422
29	414	40.3	424	35	US-09-933-524-97422
30	414	40.3	424	35	US-09-933-524A-97422
31	410.8	40.0	458	25	US-09-652-122-2749
					Sequence 1, Appl
					Sequence 26, Appl
					Sequence 26, Appl
					Sequence 3794, Ap
					Sequence 4943, Ap
					Sequence 2263, Ap
					Sequence 2853, Ap
					Sequence 521, App
					Sequence 3095, Ap
					Sequence 163, App
					Sequence 3384, A
					Sequence 2, Appl
					Sequence 2764, Ap
					Sequence 2204, Ap
					Sequence 2204, Ap
					Sequence 1233, Ap
					Sequence 5582, Ap
					Sequence 5582, Ap
					Sequence 173, App
					Sequence 46041, A
					Sequence 46041, A
					Sequence 1950, Ap
					Sequence 2930, Ap
					Sequence 1741, Ap
					Sequence 95211, A
					Sequence 95211, A
					Sequence 95211, A
					Sequence 97422, A
					Sequence 97422, A
					Sequence 97422, A

32 406.2 39.5 819 25 US-09-652-122-4862
 33 406.2 39.5 819 29 US-09-726-170-2332
 34 396.4 38.6 398 16 US-09-293-972-10989
 35 396.4 38.6 398 34 US-09-904-939-10989
 36 395.6 38.5 462 19 US-09-528-409-96481
 37 395.6 38.5 462 35 US-09-933-524-96481
 38 395.6 38.5 462 35 US-09-933-524A-96481
 39 391.4 38.1 432 19 US-09-528-409-97559
 40 391.4 38.1 432 35 US-09-933-524-97559
 41 391.4 38.1 432 35 US-09-933-524A-97559
 42 386.4 37.6 402 29 US-09-726-175-9551
 43 383.8 37.3 449 13 US-08-975-985-3212
 44 383.8 37.3 449 13 US-08-993-529-10
 45 383.8 37.3 449 34 US-09-912-292-43461

ALIGNMENTS

RESULT 1
 US-08-993-529-1
 ; Sequence 1: Application US/08993529
 ; GENERAL INFORMATION:
 ; APPLICANT: DUAN, ROXANNE
 ; APPLICANT: RUBEN, STEVEN
 ; TITLE OF INVENTION: Parotid Secretory Protein
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
 ; STREET: 9410 KEY WEST AVENUE
 ; CITY: ROCKVILLE
 ; STATE: MD
 ; COUNTRY: US
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/993,529
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BROOKES, ANDERS A.
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PF348
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 301-8439
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1028 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 49..795
 ; FEATURE:
 ; NAME/KEY: sig_peptide
 ; LOCATION: 49..100
 ; FEATURE:
 ; NAME/KEY: mat_peptide
 ; LOCATION: 103..795
 ; US-08-993-529-1

Query Match 100.0%; Score 1028; DB 13; Length 1028;
 Best Local Similarity 100.0%; Pred. No. 1.3e-279;
 Matches 1028; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CACGAGATTTTCATGAGCATCTCTCTAAACGGGTGTCAAGACAAAAGATGCTTACGCTT 60
 1 CACGAGATTTTCATGAGCATCTCTCTCTAAACGGGTGTCAAGACAAAAGATGCTTACGCTT 60
 61 TGGAACTTGTTCCT 120
 61 TGGAACTTGTTCCT 120
 121 CTGGCAATGACCTTAAGCAATGTCTGATTAAGCTGGAACCTCTCTCTCTCTCTCTCTCTCTCT 180
 121 CTGGCAATGACCTTAAGCAATGTCTGATTAAGCTGGAACCTCTCTCTCTCTCTCTCTCTCTCT 180
 181 GAGACAGTTGACAATACTCTTAAAGGCATCTTAAAGGCATCTTAAAGGCATCTTAAAGGCATCT 240
 181 GAGACAGTTGACAATACTCTTAAAGGCATCTTAAAGGCATCTTAAAGGCATCTTAAAGGCATCT 240
 241 CTTCAGAAATCCACTGCTTGGCAACTGGCCCAAGCAAGGCCAGCAAGCTGAGAATTTG 300
 241 CTTCAGAAATCCACTGCTTGGCAACTGGCCCAAGCAAGGCCAGCAAGCTGAGAATTTG 300
 301 CTGAACAATGTCATTTCTAAAGCTGCTTCCAACTAACACGGGACATTTTGGGTTGAAAAATC 360
 301 CTGAACAATGTCATTTCTAAAGCTGCTTCCAACTAACACGGGACATTTTGGGTTGAAAAATC 360
 361 AGCAACTCCCTCATCTCTGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGAT 420
 361 AGCAACTCCCTCATCTCTGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGAT 420
 421 CTGAGCTTCCCTGTACCCGCGAATGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
 421 CTGAGCTTCCCTGTACCCGCGAATGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
 481 AACCTGAAGGCTCTCTGACCTCTGACCGGAGTACAACTGAAACTGATGAACTGATGAACTGATG 540
 481 AACCTGAAGGCTCTCTGACCTCTGACCGGAGTACAACTGAAACTGATGAACTGATGAACTGATG 540
 541 CACCAGCCTGTTCGCTCTCTGCGGAGATGCGGAGATGCGGAGATGCGGAGATGCGGAGATGCGG 600
 541 CACCAGCCTGTTCGCTCTCTGCGGAGATGCGGAGATGCGGAGATGCGGAGATGCGGAGATGCGG 600
 601 TTGCTGCAACAAACACAGCCAAATCATCAACAGTTGCTGAATAGCGTGATCAACACGCTG 660
 601 TTGCTGCAACAAACACAGCCAAATCATCAACAGTTGCTGAATAGCGTGATCAACACGCTG 660
 661 AAAAGCACTGTATCT 720
 661 AAAAGCACTGTATCT 720
 721 CACTCCCTGATGTGATGTCATTCAGAGGTCGTGATGTCGATGTCGATGTCGATGTCGATGTCGAT 780
 721 CACTCCCTGATGTGATGTCATTCAGAGGTCGTGATGTCGATGTCGATGTCGATGTCGATGTCGAT 780
 781 CTGCAAAACCT 840
 781 CTGCAAAACCT 840
 841 TTCCAGTGGCTTGGCCCAACCT 900
 841 TTCCAGTGGCTTGGCCCAACCT 900
 901 CTAAACAGCTGAAAGCCTGAGTCCCAACAGAGGACCTTCCCAAGAGCTGCTGCCACCCAG 960
 901 CTAAACAGCTGAAAGCCTGAGTCCCAACAGAGGACCTTCCCAAGAGCTGCTGCCACCCAG 960
 961 ACAGTCAGAACAGCAGCTCTACACATGTTGCTCTGCCCTGGCAATTAAGGCCCATTTTC 1020
 961 ACAGTCAGAACAGCAGCTCTACACATGTTGCTCTGCCCTGGCAATTAAGGCCCATTTTC 1020
 1021 TGCAAAAA 1028
 1021 TGCAAAAA 1028


```

; OTHER INFORMATION: seq QLWKLVLICGLVLT/GT
; NAME/KEY: polyA_signal
; LOCATION: 1023..1028
; NAME/KEY: polyA_signal
; LOCATION: 1042..1058
; US-09-599-360B-26

```

```
Query Match          99.2%; Score 1019.8; DB 22; Length 1058;
Best Local Similarity 99.8%; Pred. No. 2.8e-277;
Matches 1021; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

QY	6	GATTTCATTGACGATCCCTCTCTAAACGCGTGTCAAGACAAAGATGCTTCAGCTTTGGAA	125
Db	24	gatttcagagcaaccctctctaaacgctgacaagacaaagatgcttcagcttctggaa	83
QY	66	ACTTGTCTCTGTGCGCGGTCTCACTGGGACCTCAGAGTCTCTCTTCACAATCTTGG	125
Db	84	acttgtctctctgtgcygngtgctactggyacctcagagtctctcttgacaactctgg	143
QY	126	CAATCACCTAAGCAATGCTGTGATAGCTGGAACCTGTCTTCACGAGGACTTGACAC	185
Db	144	caatgacctcaagcaatctcgggataagcggaaactgtctctcacgagggacctgagac	203
QY	186	AGTTGACAATATCTTTAAAGGCATCTCTTGAGAACTGAAGGTGCACTTAGAGTGTCTCA	245
Db	204	agtltgacaatactcttaaagcactctcttgagaaactggaaggtcgaccttaggagtgtctca	263
QY	246	GAATTCACAGTCTTTGGCACTGGCCAAAGCAGAGGCCAGGAAGCTGAGAAATTTGCTGAA	305
Db	264	gaatatcaagtcttggcaactggccaagcagaagggcccaagaagctgagaaatctgtgaa	323
QY	306	CAATGTCAATTTCTAAGCTGCTTCCTCACTACACGGGNCAATTTTGGTTGAAATTCAGCAA	365
Db	324	caatgtcatctctaaagtgtctccaactaacgcgacatcttgggttgaaaatcagcaa	383
QY	366	CTCCCTCATCTCGTATGTCAAAAGCTGAACGATCGATATGSCAAAGCCCTTAACTCTGAG	425
Db	384	ctctctcatctctggagtctcaagctgaaccgatctgagtggcaagggccttaacctgag	443
QY	426	CTTCCTGTACCCGGATCTCACTGTGCGCGGGCCCATCTATGGCCAGATTATCAACCT	485
Db	444	ctctctctgacacgcgaaatgtcactgtggccgggcccatacttggcagatctacaacct	503
QY	486	GAAAGCCCTCTTGGACCTCTGACCGCAGTCACAATTGAAACTCATCCCGACACACCA	545
Db	504	gaagagcctctctggacctctctgacctgagtcacaattgaaactgataccccagacaccca	563
QY	546	GCGTGTGCGGTCTGGGAGAAATCGCCAGTGCACCCACACGACATCTCACTTTCCCTGCT	605
Db	564	gctgtgttctgttcttggagagaaatgtgcctgagtgaccctaccagatctctcatcttctgt	623
QY	606	GGACAACACACCAAAATCATCAACAAAGTTTCGTGAAATAGCCGTATCAACAGCTGAAAAG	665
Db	624	ggacaaaacagagcaaatctatcaacaagtctcttgaaatgagctgaaatcaacacgtctgaaa	683
QY	666	CAGTGTATCTCCCTGCTGCAAGAGGAGATATGTCCACTGATCGCGATCTTCATCCACTC	725
Db	684	caactgtatctctctctgtgcagaaggagataatgtccaactgatacgaatctctatccactc	743
QY	726	CCTGGATGTGAATGTCAATTCAGCAGGTCTCGATATCTCTCAGCACAAACCCAGCTGCA	785
Db	744	ctgtgattgaaatctcatcagcagctctctgataatcctcagcacaaaacccagctgca	803
QY	786	AACCCCTCAATTTCAAGAGGACCAATGAGGAGGACCACCTGTGTTGATGTGATTGGTTCC	845
Db	804	aacctctcatctgaaaggagcgaatggggagggccactgtgtgtcattgatttggttccc	863
QY	846	ACTGGCTTGCCTCCACCCCTTATAGCATCTCCCTCCAGAAAGCTGTCTGCCACCACTAAC	905
Db	874	agtggcttgcctcccccctcttaagatctcctccaggaagctgctggcacccacataac	923
QY	906	CAGCGTGAAGCCCTAGTGTCCACACGAAAGGACCTTCCAGATACCCCTTCTCTCAAGT	965

QY	549	TGTTGCCGTCTGGGAGAAATGCCCAAGTACCCAAACAGCATCTTCCTTGTCTGGA	608
Db	552	tgtgctgctcttggagaaatgcccagttagcccaaccagcatctcaacttctctgtgga	611
QY	609	CAAAACACAGCAAAATCATCAACAAGTTTCGTGAATAGCCGTGATCAACACGCTGAAAAACAC	668
Db	612	caaacacagccaaatcatcaacaagtLcgtgaatagcgtatcaacacgctgaaaagcac	671
QY	669	TGTATCTCTCCCTGCTGCAGAAAGAGATATGTGCCACTGATCGCATCTTCATPCACTCCCT	728
Db	672	tgtatctctcttctgcagaaagagatatgtccactgatacgcgcatcttcatccactccct	731
QY	729	GGATGTGAATGTCAATTACGACGTCGTGCATAATCCTCAGCACAAAAACCCAGCTGCAAAAC	788
Db	732	ggatgtgaatgtcattcattcgcagcgtLcgtcgataatctctcagcacaaaaccagctgaaac	791
QY	789	CCTCATTTGAAATAGGACGAATCAGGAGAGACCACTGTGTGTCATGCTGATGGTTCCCACT	848
Db	792	ctctcatctgaagagacgaatgaggagagaccactgtgtgcatgctgatgtgttcccaagt	851
QY	849	GGCTTGCCCAACCCCTTTATAGCATCTCCCTCCAGGAAGCTGCTGCCACCACTTAACCAAG	908
Db	852	ggcttgcccaaccccttatagcactctcctccaggaaagtgtgtgcaaccactaaaccag	911
QY	909	CGTGAAGACCTGAGTCCCAACAGAGGACCTTCCAGATACCCTTCTCCTCACAGTCAG	968
Db	912	cgtagaagaccttagtctcccaaccaggagaccttcccagataccccttctctcaacagtcag	971
QY	969	AACAGCAGACCTCTACACATGTTGTCTCCGCCCTGGCAATAAGGCCCACTTCTGCAAAAA	1028
Db	972	aacagcagacctctacacatgtgtctgcctccctgagcaataaagagcccatcttctcaaaaa	1031

PRELIMINARY

```

RESULTS      5
US-09-652-122-4943
; Sequence 4943, Application US/09652122
; GENERAL INFORMATION:
; APPLICANT: Shy'jan, Andrew W.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.1194-001
; CURRENT APPLICATION NUMBER: US/09/652,122
; CURRENT FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/151,421
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 5020
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4943
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-652-122-4943

```

```
Query Match      99.0%; Score 1018.2; DB 25; Length 112;
Best Local Similarity 99.7%; Pred. No. 8.1e-277;
Matches 1020; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY	6	GATTTTCATGAGCATCTCTCTCTAAACCGCTGTCAAGACAAAGATGCTTCAGCTTTGGAA	65
Db	38	gatttcagagcatctctctctaaacgctgtcaagacaaaagatgttcagctttggaa	97
QY	66	ACTTTGTTCTGTGCGCGCTGCTCACTGGGACCTCAGAGTCTCTCTTCACAAATCTGG	125
Db	98	acttgttctctgtgcgctgtctcactggagccctcagatctctcttgcacaaattctgg	157
QY	126	CAATGACCTTAGCAATGTCGTGTAAGCTGTGAACCTGTTCTTCACAGGAGCTTCAGAC	185
Db	158	caaTgacctaaagaaTgcTgcgataagctggaaaccttcttcacagaggacttgagac	217
QY	186	AGTTGACAACTACTCTTAAGGCACTCTTTGAGAAACTGAAGTCCACCTAGAGCTGCTTCA	245

Db	218	agttgacataactctttaaggagccatcttgagaaactggaagglcgacctcaggagtgcttca	277
Qy	246	GAATCCAGTGTCTGGCAACTGGCCAAAGCAGAAAGCCAGGAAGCTGAGAAATTTGCTGAA	305
Db	278	gaaatccagtgcttgcaactgccaacgagaagcccgagaagctgagaaatgctgaa	337
Qy	306	CAATGTCAATTTCAAGCTGCTTCCAACTTAACACGACGACATTTTTTGGTTTGAAANTCAGCA	365
Db	338	caatgtaatttcaagctgcttccaactaacagcgacatttttggttgaaataccagca	397
Qy	366	CTCCCTCACTCTGGATGTCAAACCTGAACCGATCGATGATGCGAAAGCCTTAACCTGAG	425
Db	398	ctccctcaactctggaagtcaaaagctgaacccgatcgatggcaaaagccttaacctgag	457
Qy	426	CTTCCGTGTACCGCGAATGTCACTGTGGCCGGGCCCATCATTTGCCAGATATTCAACCT	485
Db	458	cttccctgtaaccgcgaaagtacactgtgcccggcccatactttggccagataccaacct	517
Qy	486	GAAAGCCTCCTTGGACCTCCTCAACCGCAGTCACAAATGAAACTGATCCCCACACACCA	545
Db	518	gaaagcctctcttggaacctctgaccgcagtcacaattgaaactgatcccccagacacacca	577
Qy	546	GCCTGTTCGCTCCTGGGAGAAATGGCCAGTGCACCAACACAGCATCTCACATTCTCTTGCT	605
Db	578	gcctgttgccgtctggagaatgcccgaagtgcaccaaccagcatctcaatttctcttgct	637
Qy	606	GGACAACACAGCCAAATCATCAACAAGTTCGTGAATAGCGTGATCAACACGCTGAAAG	665
Db	638	ggacaacacagccaaatcatcaacaagtctcgtgaatagcgtgatacaacacgctgaaag	697
Qy	666	CACGTGATCCTCCCTGCTCGAGAAGAGATATGTCCACTGATCGCATCTTCATCCACTC	725
Db	698	cactgatactccctgctgcagaagagatatgtccactgatcccgatcttcatccactc	757
Qy	726	CCTGGATGTGAATGTCACTTACACAGGTGTGTGAATCCTCAGACAAAAACCCAGCTGCA	785
Db	758	cttggtatgtaatgtcatctcaagcagtgctcgataatcctcagcacaaaacccagctgca	817
Qy	786	AACCCCTCATTTGAAGAGACGAATGAGGAGGACCACTGTGGTGATGCTGATTGGTTCCC	845
Db	818	aacctcaattgaagagagcgaatgaggaggaaccaactgggtgatgtgattggttccc	877
Qy	846	AGTGGCTTGGCCACCCCTTATAGATCTCCCTCCAGGAAGCTGTGCCACCACTTAAC	905
Db	878	agtggcttgcccacccctttagcatctccctccaggaagctgtgtcccaaccactaac	937
Qy	906	CACGCTGAAGAGCTGAGTCCCAACCAAGAGCACTTCCCAGATACCCCTCTCCTCACAGT	965
Db	938	caugctgaaagcctgagttcccaccagaagaccttcccagatacccccttctcctcacagt	997
Qy	966	CAGAACACAGCACTCTACACATGTTGCTTGCCCTGGCAATAAAGGCCCAATTTCTGCA	1025
Db	998	cagaaacagcgtctaacacatgttgtcctgcccctggcataaagggccatttttcag	1057
Qy	1026	AAA 1028	
Db	1058	ata 1060	

RESULT 6

```

US-09-726-170-2263
RESOLUTION 6
Sequence 2263, Application US/09726170
GENERAL INFORMATION:
APPLICANT: Gearing, David P.
APPLICANT: Holtzman, Douglas A.
TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
TITLE OF INVENTION: THEREFOR
FILE REFERENCE: 1600.2049-001
CURRENT APPLICATION NUMBER: US/09/726,170
CURRENT FILING DATE: 2000-11-29
PRIOR APPLICATION NUMBER: 60/167,858
PRIOR FILING DATE: 1999-11-29
NUMBER OF SEQ ID NOS: 2529

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2263
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-726-170-2263

Query Match 99.0%; Score 1018.2; DB 29; Length 1122;
Best Local Similarity 99.7%; Pred. No. 8.1e-277;
Matches 1020; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 6 GATTTCATGACATCCCTCTCAACCGCTGCTCAAGACAAAGATGCTTCAGCTTTGGAA 65
Db 38 gatttcacagacatccctctcaaacgctgtcaagacaaagatgcttcagcttggaa 97
QY 66 ACTTGTTCTCTGTCGGCGGCTCACTGGGACCTCAGAGTCTCTTCTTGACAAATCTTG 125
Db 98 acttgctctctgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 157
QY 126 CAATGACCTAAGCAATGCTGGATAAGCTGGAACCTGTTCTTTCAGGAGGACTTGAGAC 185
Db 158 caatgacctaaagcaatgctggataagctggaaacctgttcttcacagaggaacttgagac 217
QY 186 ACTTGACAAATCTTTAAAGCATCTTCTGAGAACTGAAGTTCGACCTAGGAGTGTCTCA 245
Db 218 agttgacaatactcttaaaaggcatctcttgagaaactgaaaggtcgacctaggagtgcttca 277
QY 246 GAAATCCAGTCTTGGCACTGGCCAGCAGACAGAGCCGACGAACTCAGAAATGCTGAA 305
Db 278 gaaatccagctcttggcaactggccagcagacagagccgaaagctgagaaatgctgaa 337
QY 306 CAATGTCATTTCTAAGCTGCTTCAACTAACAACGGACATTTTGGGTGAAATCAGCAA 365
Db 338 caatgtcatcttctaagctgcttcaactaacaacggacatcttgggttgaaatacagcaa 397
QY 366 CTCCTCATCTCGATGTCAAAGCTGAACGATCGATGATGGCAAGGCTTTAACTGAG 425
Db 398 ctccctcatctctgcatgtc aaagctgaacgctgagatgctgagatgctgagatgctgagac 457
QY 426 CTTCCTGTCACCGCGAATGTCACTGTGGCGGCGTCTCAAGTCTCTTCTTGACAAATCTTG 125
Db 458 ctctcctgtcaacgcaatgtctactgtggcgccggcccatcattggccagatttatcaacct 517
QY 486 GAAAGCCTCTCTGACCTCTGACCGAGTCACAAATGAACTGATCCCGCAGACACCA 545
Db 518 gaaagcctctctgacctctgacccgagtcacaatgaaactgatacttcacacagacacca 577
QY 546 GCCTGTTGCCCTGCTGGGGAATGCTACTGTGGCGGGCCCATCATTTGGCAGCATTTCAACCT 605
Db 578 gcctgttgccctgctggggaatgctactgtggcgggcccatcatttggcagcatcttcaacct 637
QY 606 GGACAAACACAGCAAAATCATCAACAGTTCTCAATAGGCTGATCAACACGCTGAAAG 665
Db 638 ggacaaacacagcaaaatcatcaacagttctcgaaatagcgtgatcaacacgctgaaag 697
QY 666 CACTGTATCTCTCCCTGCTGGCAGAGGAGATATGTCCACTGATCCGCTATCTTCACTCACATC 725
Db 698 caactgtatctctccctgctggcagagagatattgtccactgatccgcatcttcatcaactc 757
QY 726 CTTGGATGTGAATCTCATTGAGAGGCTGCTCGATATCTCTCAGCACAACCCAGCTGCA 785
Db 758 cttggatgtgaatctcattgagagagctgctcgatattctctcagcacaacccagctgca 817
QY 786 AACCTCATTTTGAAGAGGAGGAAATGAGGAGGACCACTGTGGTGCATGCTGATTGGTTCC 845
Db 818 aacctcatcttgaagagagagcaatgagagagcaactgctgctgcatgctgattgcttccc 877
QY 846 AGTGGCTTGCCCGACCCCTTTAGCATCTCTCCAGGAAGCTGCTGCACACCACTAAC 905
Db 878 agtggcttgcccgaccccttttagcatctctccctccaggaagctgctgcacacacacaaac 937
QY 906 CAGCGTGAAGCGCTGACTCCACACAGAGGACCTTCCCAGATACCCCTTCTCTCTCACACT 965
```

```
Db 938 cagcgtgaaagcctcagtcctccaccacgaagacaccttccagatacccttctctccacagt 957
QY 966 CAGAACACGACGCTCTACACATGTTGTCCTGCCCCCTGGCAATAAGGCCCATTTCTGCAA 1025
Db 998 cagaacacgacgacctctacacatgtgtcctgccccctggcataaaggcccaattttgacg 1057
QY 1026 AAA 1028
Db 1058 ala 1060
```

RESULT 7
US-09-726-175-2853
; Sequence 2853, Application US/09726175
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Fraser, Christopher C.
; APPLICANT: Donovan, Michael J.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: 1600.2054-001
; CURRENT APPLICATION NUMBER: US/09/726.175
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/167.859
; NUMBER OF SEQ ID NOS: 3770
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2853
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-726-175-2853

Query Match 99.0%; Score 1018.2; DB 29; Length 1122;
Best Local Similarity 99.7%; Pred. No. 8.1e-277;
Matches 1020; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 6 GATTTCATGACATCCCTCTCAACCGCTGCTCAAGACAAAGATGCTTCAGCTTTGGAA 65
Db 38 gatttcacagacatccctctcaaacgctgtcaagacaaagatgcttcagcttggaa 97
QY 66 ACTTGTTCTCTGTCGGCGGCTCACTGGGACCTCAGAGTCTCTTCTTGACAAATCTTG 125
Db 98 acttgctctctgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 157
QY 126 CAATGACCTAAGCAATGCTGGATAAGCTGGAACCTGTTCTTTCAGGAGGACTTGAGAC 185
Db 158 caatgacctaaagcaatgctggataagctggaaacctgttcttcacagaggaacttgagac 217
QY 186 ACTTGACAAATCTTTAAAGCATCTTCTGAGAACTGAAGTTCGACCTAGGAGTGTCTCA 245
Db 218 agttgacaatactcttaaaaggcatctcttgagaaactgaaaggtcgacctaggagtgcttca 277
QY 246 GAAATCCAGTCTTGGCACTGGCCAGCAGACAGAGCCGACGAACTGAGAAATGCTGAA 305
Db 278 gaaatccagctcttggcaactggccagcagagagcccgaaagctgagaaatgctgaa 337
QY 306 CAATGTCATTTCTAAGCTGCTTCAACTAACAACGGACATTTTGGGTGAAATCAGCAA 365
Db 338 caatgtcatcttctaagctgcttcaactaacaacggacatcttgggttgaaatacagcaa 397
QY 366 CTCCTCATCTCGATGTCAAAGCTGAACGATCGATGATGGCAAGGCTTTAACTGAG 425
Db 398 ctccctcatctctgcatgtc aaagctgaacgctgagatgctgagatgctgagatgctgagac 457
QY 426 CTTCCTGTCACCGCGAATGTCACTGTGGCGGCGTCTCAAGTCTCTTCTTGACAAATCTTG 125
Db 458 ctctcctgtcaacgcaatgtctactgtggcgccggcccatcattggccagatttatcaacct 517
QY 486 GAAAGCCTCTCTGACCTCTGACCGAGTCACAAATGAACTGATCCCGCAGACACCA 545
```


